Instant Notes

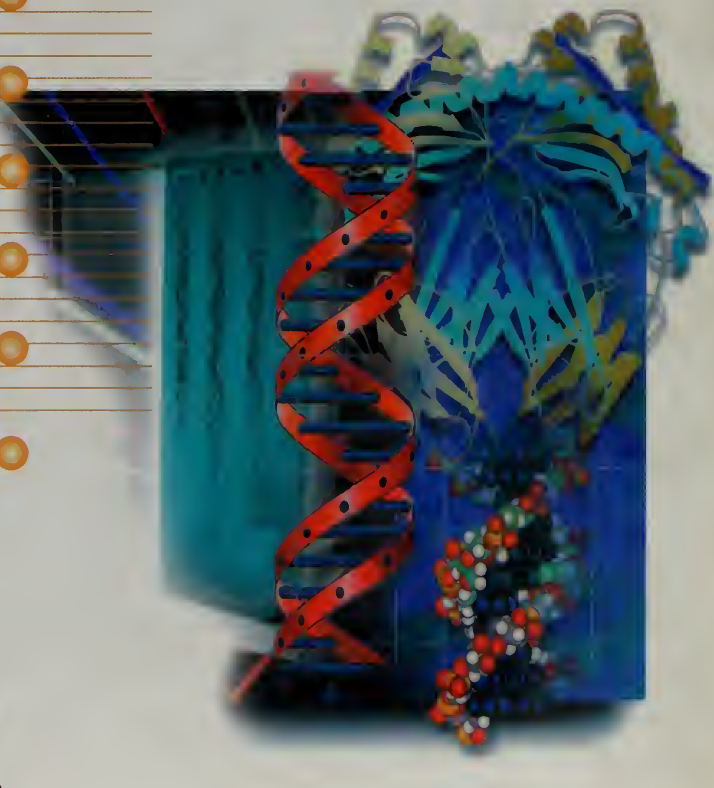
（第二版）

分子生物学

MOLECULAR BIOLOGY

现代生物学精要速览中文版





〔英〕 P. C. 特纳 A. G. 麦克伦南 A. D. 贝茨 M. R. H. 怀特 著

刘进元 李文君 王薛林 等 译校

科学出版社

BIOS SCIENTIFIC PUBLISHERS LIMITED

Digitized by the Internet Archive
in 2011 with funding from
Institute of Botany, CAS and Internet Archive

http://www.archive.org/details/fenzishengwuxue00pcte

中科院植物所图书馆
S0000193

58.178
4487

现代生物学精要速览中文版

分子生物学

(第二版)

〔英〕P. C. 特纳　A. G. 麦克伦南　A. D. 贝茨　M. R. H. 怀特　著

刘进元　李文君　王薛林等　译校

中国科学院植物研究所图书馆藏

科学出版社

北京

27522

内 容 简 介

本书是目前国外畅销的优秀教材 *Instant Notes in Molecular Biology* 第二版的中译本，由英国著名大学具丰富教学经验的一流教授编写，清华大学生物科学与技术系刘进元教授主持翻译。全书分为 19 章共 73 个主题，以简洁的形式概括了核心的分子生物学知识，既全面重点地阐述了基本理论和主要技术，又突出介绍了学科发展的前沿和动态。

本书编写风格独特、取材新颖；文字通俗易懂、简明扼要；插图简练、便于记忆；每个部分都列出要点、相关主题和进一步阅读书目，重点和主线清晰明确；书末还配有选择题利于读者复习，以取得更好的学习效果。本书专为生物学专业及生命科学相关专业的大学生设计，是指导学生快速掌握分子生物学基础知识的优秀教材；同时对授课教师制定教学计划和备课也大有帮助。

P. C. Turner, A. G. McLennan, A. D. Bates & M. R. H. White
Instant Notes in Molecular Biology
Original edition published in the United Kingdom under the title of *Instant Notes in Molecular Biology*
©BIOS Scientific Publishers Limited, 2000

图字 01-99-0185
图书在版编目（CIP）数据

分子生物学/〔英〕特纳等著；刘进元等译．-2版．-北京：科学出版社，2001.9
（现代生物学精要速览中文版）
ISBN 7-03-008383-0

Ⅰ．分… Ⅱ．①特… ②刘… Ⅲ．分子生物学 Ⅳ．Q7

中国版本图书馆 CIP 数据核字（2001）第 045753 号

责任编辑：丁海珈　谢灵玲/责任校对：陈玉凤
责任印制：刘士平/封面设计：黄华斌　高海英

科学出版社 出版
北京东黄城根北街16号
邮政编码:100717
http://www.sciencep.com
新蕾印刷厂 印刷
科学出版社发行　各地新华书店经销
*
2001年9月第　二　版　开本：787×1092　1/16
2003年2月第四次印刷　印张：23
印数：11 001—16 000　字数：514 000

定价：38.00 元

（如有印装质量问题，我社负责调换〈环伟〉）

翻译人员名单

（按章节顺序排列）

刘进元　王薛林　李文君
董晓楠　杨　涛　叶　彬
程　凌　郑　润　江　乐
林会兰　陆　韵　冯　海

丛 书 序

不几月，21 世纪即将来临，此刻，我刚完成了《生物化学》教科书第三版中所担负的写作任务，面前摆着《现代生物学精要速览》的系列书近 7 册，脑海里不时浮想生物科学的百年大事。20 世纪初（1900 年），孟德尔（Mendel）遗传学的基本定律刚被认知后不久，摩尔根（Morgan）创立了染色体基因学说。百年后的今日，整个生命科学领域的最大课题——人类基因组 DNA 全序列的测定已在进行中，包括数万个基因的 30 亿对碱基序列的解码进入 21 世纪即将宣告成功。这是一个激动人心的时代，怎不令人兴奋不已。

生命科学发生巨变，缘起于 20 世纪之初。由于数学、物理、化学广泛深刻地渗入，给现代生物科学奠定了基础，特别在 1953 年沃森（Watson）和克里克（Crick）发现了 DNA 分子双螺旋结构后，从 20 世纪 70 年代开始，分子生物学逐步形成，生命科学面目一新。近 20 年来，前沿的分子生物学和基础的生物化学出现了惊人的进展，并扩及整个生命科学，它不仅引起了学术界的极大关注，而且很大地影响了人类的生活，生命科学遂成了自然科学领域的带头学科。

生命科学进入分子水平后，才得以从本质上揭示各个层次的生命活动的真谛。当前，分子水平和细胞水平的生命科学已全面地进入了重大转变时期。以被称为龙头的人类基因组的研究为例，它的解密将使遗传、变异、生长、发展、衰老、死亡等生命现象获得认识上的飞跃。当然，基因组的解密只是解决遗传信息库的问题，每一生命活动都是由基因表达产物——蛋白质的特定群体所执行。“后基因组”研究必然落到“蛋白质组”的研究上，也就是基因组表达的全部蛋白质的整体研究。在生命科学探索的长征途中，“后基因组”的时代已经到来。

构成生命活动最重要的物质无疑是蛋白质和核酸，糖生物化学在 20 世纪 60 年代还只停留在有机化学的范畴。直到近年，才发现它在生命活动中担负着极为重要的信息功能。人体中有着 40～50 兆个细胞群，各个细胞相互黏着，细胞对底物间的相互识别，发生作用等等都依赖着细胞间的分子识别。这 40～50 兆个细胞群沿着空间坐标和时间坐标秩序井然地发生、发展构成了生命现象。担当这样复杂生命活动的识别功能，只有比核酸链、多肽链的信息量大几百、几千倍的糖链才能做到。糖链的这种功能已有很多实验证实，“糖生物学”也正兴起。无可置疑，21 世纪必将是分子生物学、生物化学共同发展，渗透到所有生命科学领域的时代。

由于分子生物学和生物化学的渗透，生命科学各个领域都发生了根本性变化，甚至古老的分类学也无例外。迄今靠电镜的形态学研究，若没有生物分子的知识甚至连文献也难看懂。细胞生物学、遗传学、神经解剖学、脑科学等等都已以崭新的内容出现。正如 19 世纪末期，近代自然科学始于物理学的革命，21 世纪自然科学的大转变，将始于生物学的革命。

生物是生物整体活动的表现，不是组成生命活动各个部分的叠加。生命科学的进展使物理学家认识到：需要变革传统物理学的世界观和方法论。眼下，生命活动的最高形式——脑活动，已成为理论物理学的最大挑战。物理学和生命科学间的传统界限也正开始改变。

数学也将在对生命科学的挑战中丰富自己的内容，发展新的学科。例如，基因组信息学，就是跨数学、逻辑学、计算机科学和生物学的学科。又如，生物拓扑学、生态几何学、脑及神经网络数学模型等等。

现代生命科学的形成是化学渗透蹬结果。今后，生命科学仍将是化学的重要结合点。几乎所有生命科学中的重大问题都将受到化学的挑战。从能量转换、生物膜、酶、生化反应机制、生物分子的结构与功能到生物大分子及其复合物等等，无一不既是生物学问题又是化学问题。

至于生命技术科学，基因工程在农业、医疗等的作用已为人所共知。生命科学与技术科学结合的深远意义在于用生命科学的原理来改造或创新工业技术。例如，对脑功能以及思维、学习、记忆、感觉等的本质的揭示必将导致计算机、人工智能等高技术领域的革命性突破。

总的看来，进入 21 世纪，生命科学自身将发生更大变革和突破，使人类越来越接近于了解生命的本质。它还将继续作为自然科学的领头学科，与其他基础学科间相互作用、相互渗透。使自然科学出现一个崭新的、繁荣的局面。

作为教育工作者，此时此刻不能不想该怎样培养以生命科学为带头的新一代。对专攻生命科学的学子，要使他们掌握最新的理论和实际；对把生命科学作为副科的各个专业的学生，也得使他们懂得生命科学的基本知识和发展概貌。要达到此目的，关键在于必得有与生命科学地位相称的教学安排，有好的课堂教学和教科书；还在于领导的英明和教师的努力。

生命科学在全面发生变革和发展。环顾各分支学科的教科书，都力求把最新成果包容进去，内容越来越广，深度越来越大，这正是现代生命科学的本来面目。在此种情况下，一些基础的内容只得从略，但篇幅仍然不断补充。学子们面对这样的大厚本，时常不知怎样在那些外围知识中找出核心的东西。去秋，在国际书展上看到一本 *Instant Notes in Biochemistry* 入手细读，发现它是一本名副其实的“精要，速览”的好书，遂向科学出版社推荐，他们经过了解研究，取得英国 BIOS Scientific Publishers Ltd. 的版权，在国内将把“Instant Notes”全系列的英文版和中文译本全部出版。我为青年学子得到出版者的慧眼识才感到十分高兴。这一套书是教科书的“新品种”，它既非大本教科书，也非简明教程。**针对学生读大本抓不住要点，读简明教程又不能得到满足的问题，它采用了创造性的格式把学生必修的学科内容，分为 70～80 个主题，用言简意赅的语言和简明、清晰的图表阐明每一重要理论或实际。既易于理解，又利于记忆。**这套新品种为年轻学生们提供了最新知识，帮助他们提高学习效率。学生们能把这样的学习书和大本的名著配合攻读，定会相得益彰，取得良好实效。我衷心愿望一代学子能迎头赶上生命科学的新时代。

王镜岩

1999 年 9 月 19 日于北京大学　朗润园

译 者 序

伴随着人类步入21世纪，分子生物学也步入了后基因组时代。基因组计划的实施不仅促进了操作DNA的分子技术的迅速发展，提供了大量的DNA内在信息，而且为分子生物学研究提出了崭新的课题。20世纪人类揭示了DNA双螺旋结构，将生物学引入分子世界，从而创建了分子生物学，随后的遗传密码的解析、基因工程技术的诞生则大大促进了分子生物学的迅猛发展。而在21世纪人类需要解决的生物学根本难题是“遗传的语言问题”，即整个基因组的调控问题，毫无疑问这也是分子生物学的根本任务。如果说DNA双螺旋结构的提出标志着分子生物学的诞生，基因工程技术的建立标志着分子生物学的重大发展，那么遗传语言问题的解决将标志着分子生物学的“成熟”，切望有志于分子生物学发展的中华学子对这一时代的早日到来做出贡献。

面对分子生物学内容多、信息量大、发展快的特点，作为在高等学府执教的教师们经常会考虑哪些基本知识、基本理论应该优先重点教授给学生；而作为主修生物科学的本科生，或者初习分子生物学的学子们又会经常面临哪些基本知识、基本理论应该优先掌握的困惑。根据多年在清华大学执教分子生物学课程的经验，觉得科学出版社引进的《现代生物学精要速览》系列之一《分子生物学》会为解决上述问题提供很好的参考。正如本书著者们在原著前言中所指出的那样，《分子生物学》力求以简练、易懂的形式传达分子生物学的精髓，以帮助读者提高学习效率。本书涉及70多个主题，基本上总括了分子生物学的基本理论、核心内容以及主要技术。在每一主题中都配有“要点”栏目，有助于抓住重点，并配有简洁的插图，便于理解。书末配有选择题以辅助读者复习，并列有大量的附加读物目录以利于大家寻找相关参考文献，去丰富知识，拓展自己的思路与视野。在科学出版社的主持下，我们翻译了该书，奉献给大家，以期帮助读者抓住重点，实现快速记忆，取得更好的学习效果。

在翻译过程中，我们力求既忠实于原文，又能符合汉语的正确表达。全书以科学出版社出版的《英汉生物学词汇》第二版规范所有术语，并对原著的某些笔误及与目前进展不符的措辞一一做了更正。

本书的主要译者是清华大学最富活力的研究生们。在第一版翻译刚完成时，又传来了该书第二版已再版的消息，为了向读者们提供最新版本，在科学出版社的精心组织下，我们又在第一版翻译稿的基础上，对第二版增加部分进行了翻译。除了在各部分末尾注明的译校者外，赵广荣、徐振彪、王泽斌、乐怡、李骥等也参加了部分工作。最后刘进元对全书进行了最终校读，以统一所用术语。由于时间紧、任务重，没有足够的时间对某些深奥的英文表达做仔细推敲，难免会有一些不确切的中文表达甚至译误，在此敬请广大读者指正，以期在再版时加以更正。

刘进元

2000年11月18日于清华园

第二版序言

为了在再版时对《现代生物学精要速览》系列之一《分子生物学》第一版做很好的修改，我们仔细研究了读者们对第一版的意见，惊喜地发现本书存在的一些疏漏以及值得引起注意的问题。由此我们所面临的挑战是如何在不改变本书现有结构的基础上，插入一定量与原内容差异很大的条目和主题，所以我们的选择是尽可能在现有主题中插入新的内容，只在绝对需要时才创建新主题。从表面上看第二版只改了很少的一部分，但实际上更新或延伸了的内容包括以下部分：蛋白质组学、LINES/SINES、信号转导、细菌人工染色体（BAC）、Z-DNA、基因枪、基因组学、DNA 指纹法、DNA 芯片、微阵列、RFLP、遗传多态性、基因组测序计划、SSCP、自动 DNA 测序、定位克隆、染色体跳查、PFGE、多重 DNA 扩增、RT-PCR、定量 PCR、PCR 筛选、PCR 诱变、简并 PCR 和转基因动物。此外还加入了三个全新的主题。毫无疑问，没有分子生物学课本会省略对克里克中心法则的讨论，在第二版中它是 D5 主题——遗传信息流的基础。另外两个迅速发展并很重要的部分是细胞周期和细胞凋亡，我们认为这两个部分都应作为主题，被分别加入到 E 的 DNA 复制部分和 S 的肿瘤病毒与癌基因部分。最后，为了保持第一版具有指导学习与辅助复习的特色，在第二版中我们根据主题顺序编入了 100 多个选择题。这一点点改进也许会大大增进本书的教学效用。

致　谢

感谢那些不厌其烦地反馈意见和建议的所有第一版读者，没有他们的建议第二版就不会有这么大的改进；感谢 BIOS 的 Will Sansom、Andrea Bosber 和 Jonathan Ray，他们不停地给我们以鼓励；最后还要感谢我们的家人，在第二版重写过程中他们又一次给予了大力的支持。

（王薛林　译　刘进元　校）

第一版序言

在刚刚过去的 20 年中，我们对遗传信息的保持、传递和表达等过程即生命本质在分子水平上的理解经历了一场革命。在许多成为这一知识大拓展的必备的基础技术进步中，占有极为重要地位的是从一种生物中分离某一特定 DNA 片段，在试管中对其进行操作，然后将其重新导入相同或不同种生物中去的能力。分子生物学正是由于重组 DNA 技术，或称遗传工程的重要贡献而得以发展。分子生物学就是要解释生物分子的结构与功能间的关系，以及这种关系是如何操纵和调控各种生化过程的，其主要目标在于 DNA、RNA 和蛋白质等大分子和大分子复合体，以及复制、转录和翻译的过程。操作这些分子的先进的实验技术是现代分子生物学的核心。分子生物学不仅可以提供这些分子的基本信息，更可以被广泛应用于开发新型且安全的产品如药物、疫苗和食品，以及遗传疾病的诊断与基因治疗。

这门学科的大发展必然导致优秀的、综合型的教科书的大量涌现。这些教科书不仅制作精美，而且对一、二年级本科生在该领域知识广度和深度的拓展都大有裨益。正是基于这一考虑，现代生物学精要速览系列之一《分子生物学》定位于以简练、易懂的形式传达该学科的精髓以有助于读者复习。这本书分为 19 个部分共 70 个主题。每个主题都有一个“要点”栏目，用极为简练的语言概括了本主题所涉及的要点。在正文中对此进行详细阐述并配有简单清楚的黑白插图。为了最好地利用这本书，必须先学习与主题相关的一些内容，要点则可用来作为快速复习的辅助。书中各主题的排列顺序合乎逻辑，而且可以从其中任何一个知识点切入阅读。正因为此，本书提供了大量的参考文献以引导读者了解相关主题。

本书的内容反映了在生命过程的分子分析中所用到的主要技术以及应用这些技术所得出的结论。它们主要以本书的作者在利物浦大学给一、二年级生物学科的本科生讲授分子生物学课程的内容为基础。A 介绍了细胞和大分子的分类并简述了用以分析的一些方法；B 着重讲述了蛋白质结构的基本要素及其结构与功能的关系；C 则讨论了 DNA 和 RNA 分子的结构及其物理化学特性，其中包括涉及超螺旋 DNA 的一些复杂概念；D 主要讲述了怎样将 DNA 整合进原核和真核生物复杂的基因组中；基因诱变、DNA 复制、DNA 重组以及 DNA 修复等相关主题则在 E 和 F 中加以阐述。

G 介绍了现有的对 DNA 进行操作的技术，简单的 DNA 克隆策略图展示了这些基本方法。如上所述，该部分巩固了我们对细胞过程分子机制的详细理解。H 描述了许多更常用的适于各种用途的克隆载体；I 着重讲述了用 DNA 文库来分离筛选新的基因序列；而 J 包含了涉及 DNA 测序和克隆序列分析方面的更加复杂且详细的方法，最后还讨论了基因克隆技术应用的某些快速进展。

原核生物基因转录的基本原理在 K 中描述，而 L 则列举了一些被细菌用来调控特定基因表达的精细机制的例子。M 和 N 讲述了与此相似但更为复杂的真核细胞转录机

制。新生 RNA 到成熟 RNA 分子的加工过程在 O 中有详细阐述，而 P 和 Q 中对这些 RNA 分子在遗传密码翻译成蛋白质序列的过程中的作用进行了描述。原核和真核生物病毒对我们理解分子信息处理所作的贡献在 R 中被详细阐述，最后的 S 对病毒的研究以及分子生物学其他领域所积累的知识是如何帮助我们深入了解主要人类疾病——癌症的发生机制做了介绍。

本书并不意在替代综合的、主流的教科书，相反希望能成为你课程笔记的一个直接补充，为你提供坚实的基础知识。大部分正文以及列在书末的进一步阅读文献中的一些论著均可为详细理解与所学课程相关的主题提供参考。而对于你们当中那些已激发了对该学科极大兴趣和热情的学生，该附加读物栏也可以指导你们阅读一些更详细和深入的文章，使你们的视野得以超越《分子生物学》的范围。不可避免地本书中会有一些遗漏，我们相信每一位读者都会从中发现不同的遗漏，但这些遗漏中的许多部分将会在现代生物学精要速览系列的其他卷，例如本书的姊妹卷《生物化学》中涉及到。

P. 特纳　A. 麦克伦南　A. 贝茨　M. 怀特

致　谢

我们首先要感谢来自家庭的支持与理解，因为我们将许许多多个原本可以与他们一起欢聚的夜晚用在了本书的起草和修改上。我们还要对我们的同事 Malcolm Bennett 和 Chris Green 在噬菌体、病毒和癌基因等章节的帮助表示感谢。我们同样要感谢 BIOS 出版社的 Jonathan Ray、Rachel Robinson、Lisa Mansell 和该系列丛书的编辑 David Hames，是他们在需要时为我们提供支持和有益的建议，并给我们以按时完成此书的适度的压力。

（李文君　译　刘进元　校）

缩略语

ADP	adenosine 5′-diphosphate	腺苷二磷酸
AIDS	acquired immune deficiency syndrome	获得性免疫缺陷综合征
AMP	adenosine 5′-monophosphate	腺苷一磷酸
ARS	autonomously replicating sequence	自主复制序列
ATP	adenosine 5′-triphosphate	腺苷三磷酸
BAC	bacterial artificial chromosome	细菌人工染色体
BER	base excision repair	碱基切除修复
bp	base pairs	碱基对
BRF	TFⅡB-related factor	TFⅡB 相关因子
BUdR	bromodeoxyuridine	5-溴脱氧核糖尿苷
bZIP	basic leucine zipper	碱性亮氨酸拉链
CDK	cyclin-dependent kinase	依赖细胞周期蛋白的激酶
cDNA	complementary DNA	互补 DNA
CHEF	contour clamped homogeneous electric field	钳位均匀电场电泳
CJD	Creutzfeld-Jakob disease	克罗伊茨费尔特-雅各布病
CRP	cAMP receptor protein	cAMP 受体蛋白
CSF-1	colony-stimulating factor-1	集落刺激因子 1
CTD	carboxyl-terminal domain	C 末端结构域
Da	Dalton	道尔顿
dNTP	deoxynucleoside triphosphate	脱氧核糖三磷酸
ddNTP	dideoxynucleoside triphosphate	双脱氧核糖三磷酸
DMS	dimethyl sulfate	二甲硫醚
DNA	deoxyribonucleic acid	脱氧核糖核酸
DNase	deoxyribonuclease	DNA 酶
DOP-PCR	degenerate oligonucleotide primer PCR	简并寡聚核苷酸引物 PCR
dsDNA	double-stranded DNA	双链 DNA
EDTA	ethylenediamine tetraacetic acid	乙二胺四乙酸
EF	elongation factor	延伸因子
ENU	ethylnitrosourea	乙基亚硝基脲
ER	endoplasmic reticulum	内质网
ESI	electrospray ionization	电喷射离子化
ETS	external transcribed spacer	外部转录间隔
FADH	reduced flavin adenine dinucleotide	还原型黄素腺嘌呤二核苷酸
FIGE	field inversion gel electrophoresis	倒转电场凝胶电泳
β-gal	β-galactosidase	β-半乳糖苷酶
GMO	genetically modified organism	基因修饰生物

GTP	guanosine 5′-triphosphate	鸟苷三磷酸
HIV	human immunodificiency virus	人类免疫缺陷病毒
HLH	helix-loop-helix	螺旋-环-螺旋（结构）
hnRNA	heterogeneous nuclear RNA	核内不均一 RNA
hnRNP	heterogeneous nuclear ribonucleoprotein	核内不均一核糖核蛋白
HSP	heat-shock protein	热激蛋白
HSV-1	herpes simplex virus-1	单纯疱疹病毒 1
ICE	interleukin-1-β-converting enzyme	白介素-1-β-转换酶
IF	initiation factor	起始因子
IHF	integration host factor	整合宿主因子
IPTG	isopropyl-β-D-thiogalactopyranoside	异丙基硫代-β-D-半乳糖苷
IS	insertion sequence	插入序列
ITS	internal transcribed spacer	内部转录间隔
JAK	Janus activated kinase	Janus 激活的激酶
kb	kilobase pairs in duplex nucleic acid，kolobases in single-stranded nucleic acid	千碱基（对）
kDa	kiloDalton	千道尔顿
LAT	latency-assiociated transcript	潜伏相关转录
LINES	long interspersed element	长散布元件
LTR	long terminal repeat	长末端重复
MALDI	matrix-associated transcript	基质相关转录
MCS	multiple cloning site	多克隆位点
MMS	methylmethane sulfonate	甲基甲磺酸
MMTV	mouse mammary tumor virus	小鼠乳腺瘤病毒
mRNA	messenger RNA	信使 RNA
NAD^+	nicotinamide adenine dinucleotide	烟酰胺腺嘌呤二核苷酸
NER	nuleotide exzcision repair	核苷酸切除修复
NLS	nulear localization signal	核定位信号
NMN	nicotinamide mononucleotide	烟酰胺单核苷酸
NMR	nuclear magnetic resonance nucleotide	核磁共振
nt	nucleotide	核苷酸
NTP	nucleoside triphosphate	核苷三磷酸
ORC	origin recognition complex	［复制］起点识别复合体
ORF	open reading frame	可读框
PAGE	polyacrylamide gel electrophoresis	聚丙烯酰胺凝胶电泳
PAP	poly（A）polymerase	poly A 聚合酶
PCNA	proliferating cell nuclear antigen	增殖细胞核抗原
PCR	polymerase chain reaction	聚合酶链反应
PDGF	platelet-derived growth factor	血小板衍生生长因子
PFGE	pulsed field gel electrophoresis	脉冲凝胶电泳
PTH	phenylthiohydantoin	乙内酰苯硫脲
RACE	rapid amplification of cDNA ends	cDNA 末端快速扩增法
RBS	ribosome-binding site	核糖体结合位点
RER	rough endoplasmic reticulum	糙面内质网

RF	replicative form	复制型
RFLP	restriction fragment length polymorphism	限制性片段长度多态性
RNA	rebonucleic acid	核糖核酸
RNA Pol Ⅰ	RNA polymerase Ⅰ	RNA 聚合酶Ⅰ
RNA Pol Ⅱ	RNA polymerase Ⅱ	RNA 聚合酶Ⅱ
RNA Pol Ⅲ	RNA polymerase Ⅲ	RNA 聚合酶Ⅲ
RNase A	ribonuclease A	RNA 酶 A
RNase H	rebonuclease H	RNA 酶 H
RNP	ribonucleoprotein	核糖核蛋白
ROS	reactive oxygen species	活性氧自由基
RP-A	replication protein A	复制蛋白 A
rRNA	ribosomal RNA	核糖体 RNA
RT	reverse transcriptase	反转录酶
RT-PCR	reverse transcriptase-polymerase chain reaction	反转录 PCR
SAM	*S*-adenosylmethionine	*S*-腺苷甲硫氨酸
SDS	sodium dodecyl sulfate	十二烷基磺酸钠
SINES	short interspersed elements	短散布元件
SL1	selectivity factor 1	选择性因子 1
snoRNP	small mucleolar RNP	核仁小核糖核蛋白
SNP	single nucleotide polymorphism	单核苷酸多态性
snRNA	small nulear RNA	核内小 RNA
snRNP	small nuclear ribonucleoprotein	核内小核糖核蛋白
SRP	signal recognition particle	信号识别颗粒
Ssb	single-stranded binding protein	单链 DNA 结合蛋白
SSCP	single stranded conformational polymorphism	单链构象多态性
ssDNA	single-stranded DNA	单链 DNA
STR	single tandem repeat	单向重复
SV40	simian virus 40	猿猴病毒 40
TAF	TBP-associated factor	TBP 相关因子
TBP	TATA-binding protein	TATA 结合蛋白
α-TIF	α-*trans*-inducing factor	α-反式诱导因子
Tris	tris (hydroxymethyl) aminomethane	三（羟甲基）氨基甲烷
tRNA	transfer RNA	转移 RNA
UBF	upstream binding factor	上游结合因子
UCE	upstream control element	上游控制元件
URE	upstream regulatory element	上游调控元件
UV	ultraviolet	紫外线
VNTR	variable number tandem repeat	可变同向重复序列
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galcatopyranoside	5-溴-4-氯-3-吲哚-β-D-半乳糖苷
XP	xeroderma pigmentosum	着色性干皮病
YAC	yeast artificial chromosome	酵母人工染色体
YEp	yeast episomal plasmid	酵母附加体质粒

目　录

丛书序
译者序
第二版序言
第一版序言
缩略语
A　细胞与大分子 …… (1)
A1　细胞分类 …… (1)
A2　亚细胞器 …… (4)
A3　生物大分子 …… (7)
A4　大分子的组装 …… (11)
B　蛋白质结构 …… (15)
B1　氨基酸 …… (15)
B2　蛋白质结构与功能 …… (18)
B3　蛋白质分析法 …… (25)
C　核酸的性质 …… (31)
C1　核酸结构 …… (31)
C2　核酸的理化特性 …… (38)
C3　核酸的光谱学和热力学特性 …… (42)
C4　DNA 超螺旋 …… (45)
D　原核与真核生物的染色体结构 …… (49)
D1　原核生物的染色体结构 …… (49)
D2　染色质结构 …… (51)
D3　真核生物的染色体结构 …… (56)
D4　基因组复杂度 …… (61)
D5　遗传信息流 …… (66)
E　DNA 复制 …… (71)
E1　DNA 复制概述 …… (71)
E2　细菌的 DNA 复制 …… (75)
E3　细胞周期 …… (79)
E4　真核生物的 DNA 复制 …… (83)
F　DNA 损伤、修复与重组 …… (87)
F1　诱变 …… (87)

F2　DNA 损伤 ………… (91)
F3　DNA 修复 ………… (94)
F4　重组 ………… (97)
G　基因操作 ………… (101)
G1　DNA 克隆概述 ………… (101)
G2　质粒 DNA 的制备 ………… (105)
G3　限制酶与电泳 ………… (108)
G4　连接、转化与重组体分析 ………… (113)
H　克隆载体 ………… (119)
H1　质粒载体的设计 ………… (119)
H2　噬菌体载体 ………… (123)
H3　黏粒、YAC 与 BAC ………… (128)
H4　真核生物载体 ………… (133)
I　基因文库与筛选 ………… (139)
I1　基因组文库 ………… (139)
I2　cDNA 文库 ………… (142)
I3　筛选流程 ………… (146)
J　克隆 DNA 的分析与应用 ………… (151)
J1　克隆的鉴定 ………… (151)
J2　核酸测序 ………… (156)
J3　聚合酶链反应 ………… (161)
J4　克隆基因的组构 ………… (166)
J5　克隆基因的诱变 ………… (170)
J6　克隆技术的应用 ………… (174)
K　原核生物的转录 ………… (179)
K1　转录的基本原则 ………… (179)
K2　大肠杆菌 RNA 聚合酶 ………… (182)
K3　大肠杆菌 σ^{70}启动子 ………… (184)
K4　转录的起始、延伸与终止 ………… (187)
L　原核生物的转录调控 ………… (193)
L1　乳糖操纵子 ………… (193)
L2　色氨酸操纵子 ………… (197)
L3　不同 σ 因子对转录的调节 ………… (201)
M　真核生物的转录 ………… (205)
M1　三种 RNA 聚合酶：性质与功能 ………… (205)
M2　RNA 聚合酶 I 基因：核糖体重复 ………… (207)
M3　RNA 聚合酶Ⅲ基因：5S 基因与 tRNA 基因的转录 ………… (211)
M4　RNA 聚合酶Ⅱ基因：启动子与增强子 ………… (215)
M5　通用转录因子与 RNA 聚合酶Ⅱ的起始 ………… (217)

N 真核生物的转录调控 ………… (221)
N1 真核生物的转录因子 ………… (221)
N2 转录调控举例 ………… (227)
O RNA 加工与核糖核蛋白复合体 ………… (233)
O1 rRNA 加工与核糖体 ………… (233)
O2 tRNA 的加工、RNA 酶 P 和核酶 ………… (239)
O3 mRNA 加工、hnRNP 和 snRNP ………… (242)
O4 可变 mRNA 加工 ………… (248)
P 遗传密码与 tRNA ………… (251)
P1 遗传密码 ………… (251)
P2 tRNA 的结构与功能 ………… (255)
Q 蛋白质合成 ………… (261)
Q1 蛋白质合成概述 ………… (261)
Q2 蛋白质合成机制 ………… (265)
Q3 真核生物蛋白质合成的起始 ………… (271)
Q4 翻译调控与翻译后加工 ………… (275)
R 噬菌体与真核生物病毒 ………… (279)
R1 病毒简介 ………… (279)
R2 噬菌体 ………… (282)
R3 DNA 病毒 ………… (287)
R4 RNA 病毒 ………… (291)
S 肿瘤病毒与癌基因 ………… (295)
S1 肿瘤病毒中的癌基因 ………… (295)
S2 癌基因的分类 ………… (299)
S3 肿瘤抑制基因 ………… (302)
S4 凋亡 ………… (306)
进一步阅读文献 ………… (311)
选择题 ………… (315)
答案 ………… (331)
索引 ………… (333)

A1 细胞分类

要 点

真细菌	真细菌（eubacteria）具原核生物（prokaryote）的构造。每个真细菌都有质膜（plasma membrane），通常其膜外包裹一层刚性（rigid）的细胞壁（cell wall），其内没有分隔小室（intracellular compartment），有一条主要的环状染色体。真细菌可以是单细胞或多细胞。大肠杆菌（*E. coli*）是研究得最深入的一种真细菌。
古细菌	古细菌（archaea）在结构上与原核生物相似。但是古细菌很可能是从真核生物（eukaryote）分支出来的，这发生在它们共同的祖先与真细菌歧化之后。古细菌常生活于极端环境。从生化角度上看，古细菌在某些方面接近真细菌，而在另一些方面则接近真核生物，也存在一些自己独特的生化特性。
真核生物	植物、动物、真菌（fungi）和原生生物（protist）细胞拥有由脂膜（lipid membrane）包被的亚细胞结构如细胞核（nuclei）、线粒体（mitochondria）和内质网（endoplasmic reticulum）等。这些细胞器是各种独特的生化反应进行的场所，也是真核生物的重要特征。
分化	大多数多细胞真核生物，在发育过程中细胞群经历分化形成具特殊功能的器官（例如：肝、脑和肾）。在大多数情况下，这些器官拥有相同的DNA，但转录不同的基因。正像所有其他细胞发育过程一样，分化由基因控制。不同功能细胞间的协调需要细胞间的信息交换。
相关主题	亚细胞器（A2）　　噬菌体与真核生物病毒（R） 原核与真核生物的染色体结构（D）

真细菌

真细菌是**原核生物**二个亚门之一。原核生物是最简单的活细胞，通常直径1～10μm，存在于从动物内脏到酸性温泉等各种适宜环境中。分类上，可根据其组构来鉴别（图A1.1）。原核生物由嵌有蛋白质的脂双分子层组成的**细胞膜**包被，膜上蛋白质可以允许小分子出入。大多数原核生物在质膜外还含有刚性的细胞壁，以避免细胞在渗透压与胞内显著不同的环境中膨胀或收缩。胞内（**细胞质**）通常含有一个折叠成**类核体**附着在质膜上的环状染色体（参见D1）、携有一定遗传信息的质粒［小分子的脱氧核糖核酸（DNA），参见G2］、核糖核酸（RNA）、核糖体（合成蛋白质的

场所，参见 Q）以及执行细胞内代谢反应的蛋白质。这些蛋白质中的部分与质膜相连，但是不像**真核生物**有明显的**亚细胞器**将细胞分成不同的代谢小室。原核生物表面可能带有**纤毛**和**鞭毛**，纤毛可使细胞能够附着在其他细胞的表面上，而鞭毛的旋转运动可使细胞游动。大多数原核生物是单细胞的，但是有些具有多细胞结构，其中某些细胞行使特殊的功能。真细菌与**古细菌**的主要区别在于其生化性质上。真细菌大肠杆菌（*Escherliclia coli*）有一个 4600kb 的**基因组**，携有足以编码约 3000 种蛋白质的遗传信息。大肠杆菌的分子生物学研究进行得非常深入。最简单的生殖道支原体（*Mycoplasma genitalium*）的基因组仅由 580kb 的 DNA 组成，仅编码 470 种蛋白质，仅有有限的代谢容量。

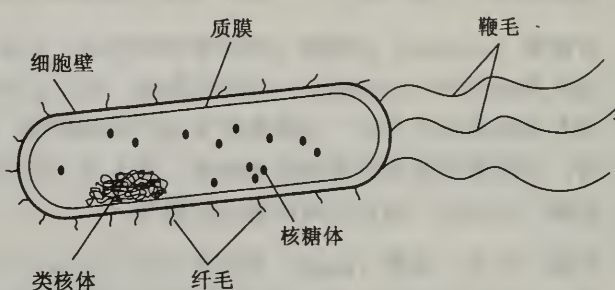


图 A1.1　典型的原核细胞结构图

古细菌

古细菌是原核生物的第二个亚门，常生活于极端环境中。从结构上看，古细菌与真细菌相似。然而根据核糖体 RNA（rRNA）的分子进化，古细菌与真细菌差别就如同原核生物与真核生物差别一样，古细菌具有一些独特的生化性质，例如膜脂由醚键连接而不是酯键（参见 A3）。古细菌甲烷球菌（*Methanococcus jannaschii*）基因组有 1740kb，编码 1738 种蛋白质。基因组间的比较表明，在能量产生与新陈代谢方面，古细菌同真细菌有着许多相同之处，而复制、转录和翻译则更接近真核生物。似乎古细菌与真核生物拥有一个由真细菌的祖先歧化而来的共同祖先。

真核生物

真核生物分为四个门，包括**动物**、**植物**、**真菌**和**原生生物**（藻类和原生动物）。在结构上，真核生物具有膜包围（图 A1.2）的行使特殊代谢功能的细胞器（参见 A2），其体积要比原核细胞大，直径约 10～100μm。真核生物有质膜包围，质膜高度皱折以增加其表面积。植物、许多真菌和原生生物也有一个刚性细胞壁。细胞质是一个高度有序的胶状物，其中含有细胞器、核糖体和被称为**细胞骨架**能控制细胞形状和运动并组织代谢功能的一系列蛋白质纤维。这些纤维包含由**微管蛋白**组成的**微管**和由**肌动蛋白**组成的**微丝**（参见 A4）。许多真核生物是多细胞的，在发育过程中不同群细胞经**分化**形成机体的特定组织。

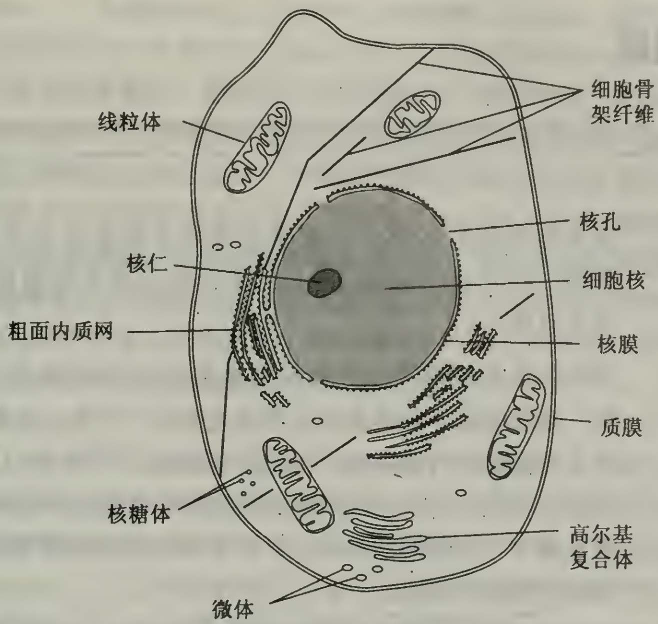


图 A1.2 典型的真核细胞结构图

分化

细胞分裂时，子细胞几乎与亲本细胞完全一样，或者通过改变基因表达模式变成在功能上与亲本细胞不一样的细胞。

在原核生物和较低等的真核生物中，孢子形成是这类**细胞分化**的一个例子（参见 L3）。在复杂的多细胞真核生物中，胚胎细胞高度分化成特定细胞，如肌肉、神经、肝和肾。除了几种例外的情况，在所有这些细胞中 DNA 的含量保持不变，而是被转录基因群体发生了改变。分化是由控制发育的基因来调控的（参见 N2）。这些基因的突变导致畸变，如果蝇（*Drosophila*）在本该长触角的部位长出了腿。对这些突变基因的研究有助于对胚胎发育的了解。在多细胞生物中，不同组织和器官间的活动协调是通过它们之间的信息来控制的，包括各种信号分子如神经传递介质、激素和生长因子。这些分子由某一种组织分泌，而通过细胞表面的特异受体作用于另一组织。

A2　亚细胞器

要　点

细胞核	在多染色体的细胞中，有膜包围的细胞核包含了大量的细胞DNA，DNA的转录和RNA的加工均发生在细胞核内。细胞核内含有核仁。
线粒体与叶绿体	线粒体是细胞呼吸、将养料氧化为CO_2和水并产生ATP的场所。线粒体由共生的原核生物演化而来，尽管大多数线粒体蛋白质是在细胞核内编码的，但它们仍然包含一些DNA、RNA和蛋白质合成机构。光合作用发生在植物和藻类的叶绿体中。叶绿体除了拥有一个含捕光色素——叶绿素的类囊体膜系统外，与线粒体的结构基本相似。
内质网	光面内质网是一个进行脂类生物合成和生物异源物质代谢反应的细胞质膜体系。粗面内质网附着在核糖体上，进行膜蛋白和分泌蛋白的合成，这些合成蛋白由小泡运至高尔基体做进一步加工与分检。
微　　体	溶酶体含有降解性水解酶；过氧化物酶体含有破坏某些有潜在危害的自由基和过氧化氢的酶；植物的乙醛酸循环体执行乙醛酸循环反应。
细胞器的分离	打碎质膜后，就可以通过差速离心和密度梯度离心（既有速度区带，又有等密度区带）相结合的方法将各亚细胞器分离纯化。纯度可由检测细胞器特异性酶来判定。
相关主题	细胞分类（A1）　翻译调控与翻译后加工（Q4） rRNA加工与核糖体（O1）

细胞核

在多染色体细胞中，真核生物的细胞核携带细胞的遗传信息，每条染色体含有一个DNA分子（参见D2、D3）。细胞核由一脂双分子层即核膜包围。核膜上有允许适度大小分子通过的孔（参见图A1.2）。RNA转录发生在（参见M）细胞核中，加工后的RNA分子（参见O）进入细胞质并在那里翻译（参见Q）。**核仁**是细胞核内的小体，是rRNA合成和核糖体进行部分组装的场所（参见M2与O1）。

线粒体与叶绿体

线粒体是细胞呼吸即营养物质氧化以ATP形式产生能量的场所。其直径约1～2μm，每个细胞中含有1000～2000个。**线粒体**有一个光滑的外

膜和一个折叠的内膜，内膜突出折叠成**嵴**（cristae）（参见 A1，图 A1.2），其内含有一个小的环状 DNA 分子、线粒体特异 RNA 和可合成一些线粒体蛋白的核糖体。然而大部分的线粒体（和叶绿体）的蛋白均是由细胞核 DNA 编码并在细胞质中合成。这些蛋白均具有特定的**信号序列**引导它们进入线粒体（参见 Q4）。植物的**叶绿体**是进行光合作用的场所，即依靠光同化 CO_2 与水形成碳水化合物和氧气。虽然叶绿体比线粒体稍大，但拥有相似的结构，只是对应于线粒体的嵴，叶绿体在内膜腔内有第三膜体系（**类囊体**）。类囊体上含有能捕捉光能进行光合作用的叶绿素。叶绿体遗传上也是部分独立于细胞核。据说线粒体与叶绿体都是由曾与原始具核真核生物有过共生关系的原核生物进化而来。

内质网

内质网是细胞质内延伸的膜体系，与核膜相连（参见 A1，图 A1.2）。在大多数细胞中可以见到两种形式：光面内质网和粗面内质网。**光面内质网**携带有许多膜结合的酶，包括那些相关于脂类生物合成、**外来物质**（xenobiotic）如药物的氧化与解毒的酶。**粗面内质网**之所以被这样称呼，是因为其上有许多核糖体的存在。核糖体特异合成由细胞分泌的蛋白质，如胞质蛋白或乳蛋白或者那些被定位于质膜或某些细胞器上的蛋白质。不同于那些早期进入粗面内质网膜的质膜蛋白，这些蛋白质被转运到粗面内质网**内腔**并且在那里被修饰，通常是**糖基化**（参见 Q4）。在粗面内质网上合成的脂类和蛋白质以特定的**转运小泡**转运至具有一迭扁平膜泡的**高尔基体**，在那里进一步被修饰、分检并引至其最终目的地（参见图 A1.2）。

微体

溶酶体是小的有膜包围的细膜器，从高尔基体出芽而成，并且含有各类能降解蛋白质、核酸、脂类和糖类的消化酶。溶酶体起着从细胞外带进大分子或来自损伤细胞器的大分子的回收中心的作用。一些产生高活性自由基和氢过氧化物的代谢反应被限制在被称作为**过氧化物酶体**的细胞器中，以防止这些分子损伤细胞组分。过氧化物酶体拥有可以破坏过氧化氢的催化酶：

$$2H_2O_2 \longrightarrow 2H_2O + O_2$$

乙醛酸循环体是特定的植物过氧化物酶体，进行乙醛酸循环。溶酶体、过氧化物酶体和乙醛酸循环体合称**微体**。

细胞器的分离

真核细胞的质膜可以用各种方式加以破坏，包括渗透压冲击、可控制的机械剪切和某些非离子去污剂作用。大小和密度不同的细胞器，例如细胞核与线粒体、可以根据它们的**沉降系数**值不同由**差速离心法**互相分离并与其他细胞器分开（参见 A4）。**细胞裂解物**以一个足以仅沉淀最重细胞器，通常即细胞核的速度离心来分离。含有各种其他细胞器的上层悬液被移去并在更高的速度下进行离心以沉降线粒体，依次可分离其他细胞器（图 A2.1a）。这项技术也被用于分离含不同大小细胞的悬液，如血液中的

红细胞、白细胞和血小板。在这些细胞的粗提物中，细胞核与线粒体通常需要用**密度梯度离心**方法作进一步纯化。这种方法也被用来分离密度相似的细胞器。在**速度区带离心**中，混合物被加入离心管中预先铺好的适当介质的浓度（形成密度）梯度顶层。通过离心，不同成分将根据他们的沉降系数以不同的速度下沉并形成相互分离的带或区带（图 A2.1b)。建立介质密度梯度的目的是为了阻碍分离后组分的扩散混合（即提供稳定性），并且确保组分的线性分离速率（当组分进一步向离心管下部运动时，它抵消组分的加速)。在**平衡（等密度）离心**，密度梯度延伸至比混合物中的一个或多个组分大一些的密度，因此这些组分在等于他们自己密度的点达到平衡并且停止移动。在这种情况下，密度梯度可以是预先建立，然后将样品加在上层，也可以是样品与梯度物质混合在一起离心时自动形成（图 A2.1c)。密度梯度用下列物质建立：蔗糖、Ficoll（一种合成多糖)、metrizamide（一种合成的重碘化合物）或者 CsCl（参见 C2 和 G2)。亚细胞组分的纯度可以用电子显微镜或测定与特定器官相关的已知酶活来判定，例如可用琥珀酸脱氢酶来判定线粒体。

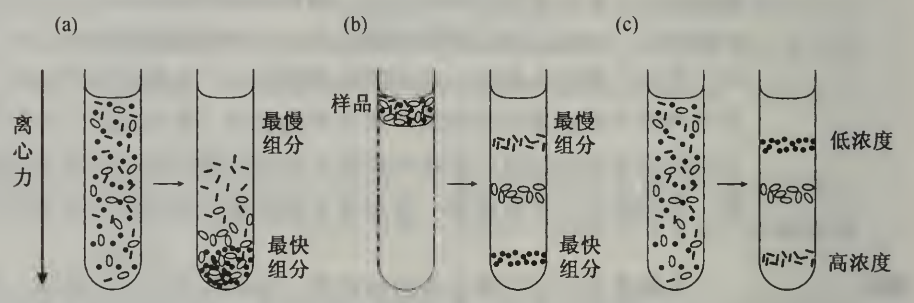


图 A2.1 离心技术。(a) 差速离心；(b) 速度区带离心和 (c) 等密度离心

A3 生物大分子

要 点

蛋白质和核酸	蛋白质是氨基酸聚合体，核酸（DNA 和 RNA）是核苷酸的聚合体，都是贮存和表达遗传信息所必需的组分。除此以外，蛋白质还有许多额外的结构与功能作用。
多 糖	α-直链淀粉与纤维素是葡萄糖分别以 α（1→4）和 β（1→4）糖苷键相连而成的聚合体。淀粉是植物体内葡萄糖的贮存形式，不仅含有 α（1→4）直链，还含有 α（1→6）支链。纤维素在植物体内形成结构纤维，而在动物体内葡萄糖以糖原形式贮存。几丁质是 *N*-乙酰氨基葡糖的聚合体，存在于真菌细胞壁和节肢动物的外骨骼中。黏多糖（mucopolysaccharide）是结缔组织的重要组分。
脂 类	含有饱和与不饱和脂肪酸的甘油三酯，分别是动物和植物体内的主要贮存脂类。两类脂肪酸的结构差异导致了动物体内的甘油三酯呈固体，而植物体内甘油三酯呈液体。磷脂与鞘脂除含有脂肪酸成分外还含有极性基团，是所有细胞膜的重要组分。
复杂大分子	核蛋白含有核酸和蛋白质，例如端粒酶（telomerase）与核糖核酸酶 P（ribonuclease P）。糖蛋白与蛋白多糖（mucoprotein）是以共价键与糖类相连的蛋白质，通常存在于细胞外表及细胞外间隙。与脂相连的蛋白质和脂蛋白中脂类与蛋白质的相连有共价或非共价两种形式。糖脂含有脂与糖两种组分。这些混合大分子复合物提供了比其组分更多的功能。
相关主题	大分子的组装（A4） 核酸的性质（C） 蛋白质结构（B）

蛋白质和核酸

蛋白质是氨基酸以肽键共价连接而成的聚合体。氨基酸与蛋白质的详细结构将在 B 中叙述。蛋白质起结构和功能双重作用。核酸（DNA 和 RNA）是核苷酸的聚合体，由含氮碱基、五碳糖和磷酸组成。核苷酸与核酸的详细结构将在 C 中详细介绍。核酸参与遗传信息的贮存与加工，但这些信息的表达则需要蛋白质。

多糖

多糖是单糖以糖苷键共价连接成的聚合体，其功能主要作为营养糖源或结构组分。**纤维素**和**淀粉**在植物体内含量丰富，两者都是葡萄糖多聚

体，只是在葡萄糖单体的连接方式上不同。纤维素是以 β（1→4）糖苷键相连成的线性多聚体（图 A3.1 a），是植物细胞壁的主要结构成分，大约 40 条平行链形成水平片层而相互堆积，链与片层之间以氢键相连（参见 A4）形成坚固的、不溶的纤维。淀粉是糖的贮存形式，常常存在于细胞内的大颗粒中，可迅速水解释放出葡萄糖以供新陈代谢所需。淀粉含有两个组分：直链淀粉，以 α（1→4）糖苷键相连的线性多聚体（图 A3.1b）；支链淀粉，除含有 α（1→4）糖苷键相连外，还含有 α（1→6）相连的分枝。支链淀粉可聚合达 10^6 个葡萄糖残基，是已知最大的分子之一。淀粉中不同的连接形成了难以紧密堆积的卷曲构型，造成淀粉可溶于水的特性。真菌和一些动物组织（例如肝和肌肉）以**糖原**的形式贮存葡萄糖，一种像支链淀粉的分枝多聚体。**几丁质**存在于真菌细胞壁、昆虫及甲壳虫的外壳中。几丁质像纤维素，只是其单体为 *N*-乙酰氨基葡糖。**黏多糖**形成分散有结缔组织丝状蛋白的凝胶状溶液。大分子多糖的结构确定是很复杂的，因其大小和组成各不相同，不能用研究核酸和蛋白质时所用的遗传工程技术来研究多糖。

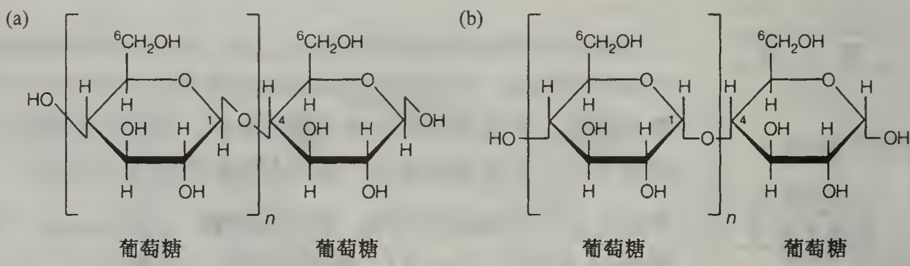


图 A3.1　纤维素（a）［β（1→4）连接］和（b）直链淀粉［α（1→4）连接］的结构图，在图中已标出了碳原子 1，4 和 6 位。若以 α（1→6）相连会产生支链淀粉以及糖原中的分枝

脂类

尽管将单一脂类称为大分子并不合适，但是许多脂类是由较小的单体构成，并参与许多大分子装配（参见 A4）。自然界中较大的脂类分子主要是碳氢化合物，难溶于水。一些脂类相关于能量的贮存和转移，而另一些则是膜、保护性外鞘以及其他细胞结构的关键组分。**甘油酯**是由一个、两个或三个长链脂肪酸与一分子甘油以酯键相连而成的。动物的甘油三酯内，都是饱和脂肪酸，不含双键线性链，分子可以紧密堆积在一起，所形成的脂肪酸脂肪是固体。植物油组成中含有一个或更多个双键的不饱和脂肪酸，其链的夹角阻碍了脂肪酸的紧密堆积，因此在室温下植物油呈液体。膜中含有**磷脂**，磷脂由两分子脂肪酸和一分子磷酸以酯键与甘油相结合，其磷酸通常也与一个小分子如丝氨酸、乙醇胺、肌醇、胆碱以酯键相连（图 A3.2）。膜中还含有**鞘脂**，如**神经酰胺二糖**，其中的长链氨基乙醇鞘胺醇含有由酰胺键相连的脂肪酸。磷酸胆碱与脑胺相连形成**鞘磷脂**。

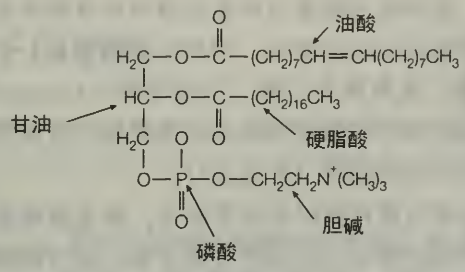


图 A3.2 一个典型的磷脂：含有酯化硬脂酸和油酸的磷脂酰胆碱

复杂大分子

许多大分子含有一种以上的生物大分子以共价或非共价键相结合。这可大大地增加所形成的复合分子的功能或结构容量。例如几乎所有的酶都是蛋白质，但某些蛋白质以非共价键与对催化活性极为重要的 RNA 组分相结合。核酸与蛋白质结合体被称为**核蛋白**，例如：**端粒酶**，一个真核染色体末端的复制酶（参见 D3 和 E3）；**核糖核酸酶 P**，一个催化转移 RNA（tRNA）成熟的酶。在端粒酶中 RNA 可作为端粒 DNA 合成模板而起作用，而在核糖核酸酶 P 中，RNA 含有酶的催化位点。核糖核酸酶 P 是一个**核酶**（参见 O2）。

糖蛋白既含有蛋白质又含有糖（其比例可占重量＜1%，甚至＞90%），糖基化是蛋白质翻译后修饰的通常形式（参见 Q4）。糖总是与蛋白质的表面共价结合，而不是在内部，且常在组成上有变化，这就造成了糖蛋白的微不均一性（图 A3.3），同时这也增加了糖蛋白研究的困难。糖蛋白的功能可涵盖整个蛋白质活性范围，并通常在细胞外还有活性。糖蛋白是细胞膜的重要组分，并介导细胞与细胞间的识别。

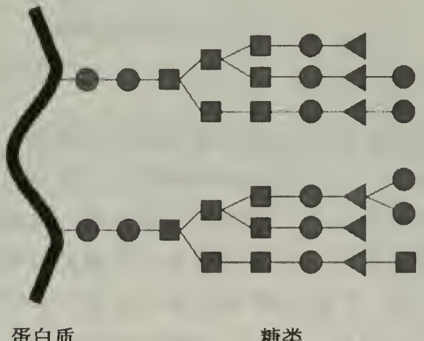


图 A3.3 糖蛋白结构。不同的符号代表不同的单糖单位（例如：半乳糖，*N*-乙酰葡糖氨基）

蛋白多糖是蛋白质和黏多糖组成的大分子复合物（＞10^7Da），存在于细菌细胞壁和结缔组织细胞间隙，其糖单元常常含有硫酸基团，这使得糖蛋白高度亲水。高度水化再加上它们的长度（＞1000 单位）造成了蛋白

多糖溶液的高黏度，在细胞间隙蛋白多糖起润滑剂及缓冲刺激的作用。

脂连接蛋白与脂类组分共价结合，脂类组分通常是一个脂肪酰基（如豆蔻酰基或棕榈酰基）或类异戊二烯［如法尼基（farnesyl）或牻牛基牻牛基］基团。这些基团通过与膜脂的疏水作用把蛋白质固定在膜中，同时也促进了蛋白质与蛋白质的结合（参见 A4)。

在**脂蛋白**中，脂类与蛋白质以非共价结合，因为脂很难溶于水，脂类在血液中以脂蛋白的形式转运。这基本上是甘油三酯和胆固醇的颗粒，并被一层由磷脂、胆固醇和蛋白质（脱脂蛋白质）构成的鞘所包被。脱脂蛋白质的结构是疏水氨基酸向颗粒的内部，而带电的以及极性氨基酸（参见 B1）朝向颗粒外部的水环境，因此该颗粒是可溶于水的。

糖脂，包括脑苷脂和神经节苷脂，脂与糖以共价键相结合，在脑细胞和神经细胞的膜中含量特别丰富。

A4 大分子的组装

要点

蛋白质复合体 真核生物细胞骨架由各种蛋白质复合体构成。蛋白质复合体包括微管（由微管蛋白构成）、微丝（由肌动蛋白与肌球蛋白构成）和中间纤维（由多种蛋白质构成）。这些蛋白质复合体组构细胞的形状、细胞运动和亚细胞器。纤毛与鞭毛也是由微管、动力蛋白（dynein）以及微管连接蛋白（nexin）复合而成。

核蛋白 细菌 70S 核糖体由一个 50S 大亚基和一个 30S 小亚基组成。前者含有 23S 与 5S RNA 分子与 31 种蛋白质，后者包含一个 16S RNA 分子与 21 种蛋白质。真核生物的 80S 核糖体含有 60S（28S、5.8S 和多种 5S RNA）和 40S（18S RNA ）两个亚基。染色质含有 DNA 和碱性组蛋白。病毒也是一种蛋白质复合体。

膜 膜磷脂和鞘脂形成极性基团在外表、烃链在内部的双分子层。膜蛋白可能是外周蛋白或整合蛋白，并起着受体、酶、转运蛋白或细胞间介质的作用。

非共价相互作用 大量的弱相互作用维系大分子组装在一起。电荷与电荷、电荷与偶极以及偶极与偶极之间的相互作用构成了全部或部分带电原子间的吸引力。氢键和排斥水的疏水作用也很重要。

相关主题

生物大分子（A3） rRNA 加工与核糖体（O1）
染色质结构（D2） 噬菌体与真核生物病毒（R）

蛋白质复合体 细胞结构包含着许多由同种或不同种生物大分子组成的大分子复合体。细胞的主要构件和运动元件中许多是由蛋白质复合体构成的。**细胞骨架**就是一系列蛋白质微丝，这些微丝组构细胞的形态、控制着细胞的运动以及亚细胞器在细胞内的分布。**微管蛋白**是一个 110kDa 的球形蛋白，其纵向排列呈长多聚体即形成微管（图 A4.1，参见 B2)。微管是细胞骨架真核生物**纤毛**和**鞭毛**以及许多细胞表面毛状结构的主要组分。这些结构物的运动可以移动细胞或移动液体穿过细胞表面。纤毛也含有动力蛋白（dynein）和微管连接蛋白（nexin）。

由**肌动蛋白**组成的**微丝**与**肌球蛋白**一起形成收缩装置，可引起细胞质运动。肌动蛋白和肌球蛋白也是肌肉纤维的主要组分。细胞骨架中的**中间纤维**由包括**角蛋白**在内的各种蛋白质组成，并具有包括巩固细胞结构等多种功能。在所有情况下，运动所需能量是由 ATP 或 GTP 的偶联水解所提

供。

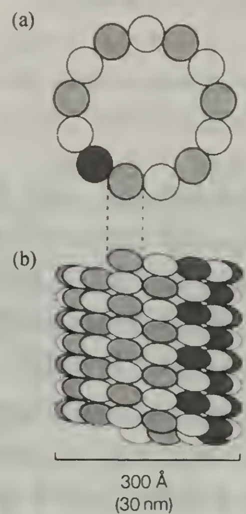


图 A4.1 微管中微管蛋白 α 和 β 亚基的横截面（a）和表面（b）的示意图（参见 B2）经 Freeman，W. H. 及其出版商同意，引自由 Lubert Stryer 主编的《生物化学》Stryer，1995

核蛋白

核蛋白由核酸与蛋白质组成。**核糖体**是较大的细胞质核糖核酸蛋白质复合体，是蛋白质的合成场所（参见 Q）。细菌 70S 核糖体含有大亚基（50S）和小亚基（30S），总分子质量为 2.5×10^6Da（50S 的 **S 值**是**沉降系数**的数值 S，它描述了大分子或粒子在离心场中沉降的速率。由分子或颗粒的质量与形状决定，因此 S 值不是可加的）。50S 亚基包含 23S、5S RNA 分子和 31 种不同蛋白质，而 30S 亚基包含 16S RNA 分子和 21 种蛋白质。真核生物的 80S 核糖体含有 60S（包括 28S、5.8S 和多种 5S RNA）和 40S（含 18S RNA）两个亚基。在合适的条件下，rRNA 和蛋白质的混合物在体外可以精确的程序自动组装成有功能的核糖体。可见核糖体结构的全部信息都贮藏在其组分的结构中。RNA 不仅是核糖体蛋白组装的框架，还参与信使 RNA 的结合与肽链合成的催化（参见 Q2）。

染色质是构成真核生物染色体的元件，是一个由大约等量的 DNA 与小分子的碱性蛋白质即称为组蛋白（参见 D1）所组成的**脱氧核蛋白**复合体。这些复合体形成被称为**核小体**的重复单位。核小体和许多其他蛋白质复合体的正确组装均需要装配蛋白或**伴娘蛋白**。组蛋白中和了 DNA 糖-磷酸骨架负电荷间的排斥力，并使 DNA 在染色体中紧密堆积在一起。**病毒**是核蛋白复合体的另一个例子，将在 R 中被详述。

膜

将磷脂与鞘脂置于水中，会自发地形成一个极性基团在外、非极性烃链在内的**脂双分子层**，这是所有生物膜的结构基础。这样的膜构成了细胞

和细胞器的边界，并且不带电荷的分子可选择性透过。精细的脂类成分随细胞与细胞器的不同而异。蛋白质也是细胞膜的一个主要组分（图 A4.2）。**外围膜蛋白**疏松地结合在外表或通过一个脂或**糖基磷脂醇**锚定在膜上，且易于流动。**镶嵌膜蛋白**被埋在膜中，不破坏膜就不能移动这些蛋白。当**跨膜蛋白**完全横跨双层膜时，一些蛋白质则从膜的内外表面突出（portrude），这些蛋白质既有处于细胞膜外的结构域又有细胞内**结构域**（参见 B2）。这些蛋白质的跨膜区主要含有疏水性氨基酸。膜蛋白有多种多样的功能，例如：

- 作为激素和神经递质等信号分子的受体；
- 作为在吸收物质前降解细胞外分子所需的酶；
- 作为选择性转运小分子、极性离子和分子的孔或通道；
- 作为细胞与细胞相互作用的介质（主要是糖蛋白）。

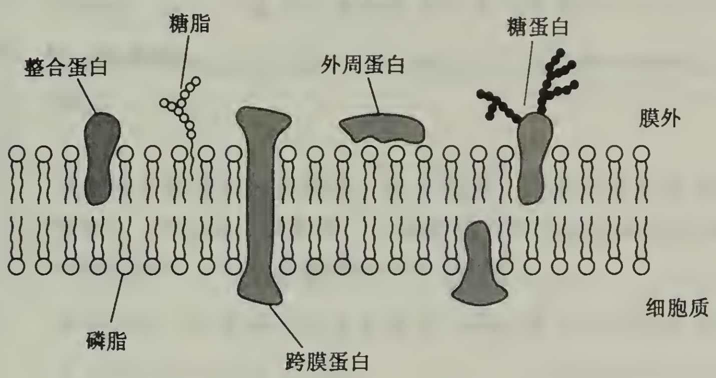


图 A4.2 显示主要大分子组分的细胞膜结构示意图

非共价相互作用

大多数大分子的组装是由大量不同的非共价相互作用维系在一起的。**电荷-电荷**相互作用（**盐桥**）存在于生理 pH 条件下带有相反电荷的离子基团之间，例如在 DNA 中的带负电的磷酸基团和带正电的 DNA 结合蛋白诸如组蛋白的赖氨酸和精氨酸侧链之间（参见 D2）。**电荷-偶极**和**偶极-偶极**间相互作用比较弱，且由于分子中的电荷的不对称分布，当其中之一或两个都是偶极时，才形成这种较弱的相互作用（图 A4.3a）。甚至像甲基这样不带电的基团由电子运动引起的瞬时偶极的作用也可以有弱的相互吸引（**色散力**）。

电中性分子间的非共价结合被总称为**范德华力**。**氢键**也是重要的一类，在供体基团的一个共价结合的氢原子（例如： —O—H 或 —N—H ）和受体基团上一对非成键电子（例如： O═C— 或 N— ）之间形成（图 A4.3b）。氢键和包含偶极的其他相互作用是具有方向性的，因此这有助于确定大分子形状和分子间相互作用的特性。不带电和非极性物质的存在，例如脂类，在水环境中会在水分子中建立一个高度有序的结构。当体系的熵值减小时，这在能量上是不利的。因此非极性分子倾向于

聚集起来以减小暴露给水的表面积。这种引力被称为**疏水相互作用**，是蛋白质-蛋白质和蛋白质-脂类相互作用以及核酸结构中的主要稳定力。

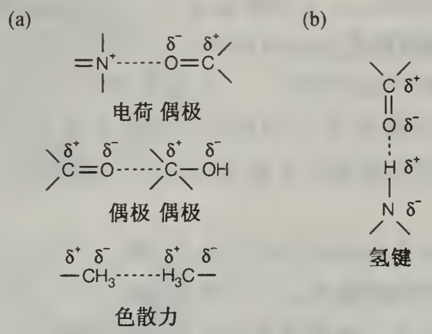


图 A4.3 范德华力（a）和氢键（b）的例子

（董晓楠 译 刘进元 校）

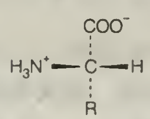
B1 氨基酸

要 点

结 构	蛋白质中已发现的20种常见的氨基酸都有一个与质子、氨基、羧基相连的手性α-碳原子和具有不同的理化特性的侧链。氨基酸在溶液中表现为两性离子，蛋白质中的稀有氨基酸通过翻译后修饰产生。
带电荷的侧链	谷氨酸与天冬氨酸都含一个羧基，通常给蛋白质增加一个负电荷。赖氨酸有一个ε-氨基、精氨酸有一个胍基、组氨酸有一个咪唑基，这三个碱性氨基酸通常给蛋白质增加一个正电荷。
不带电的极性侧链	丝氨酸和苏氨酸有羟基、天冬酰胺与谷氨酰胺有亚胺基、半胱氨酸有一个巯基。
非极性脂肪烃侧链	甘氨酸是最简单的氨基酸，没有侧链。脯氨酸是次级氨基酸（亚氨酸），丙氨酸、缬氨酸、亮氨酸和异亮氨酸有疏水的脂肪烃基。甲硫氨酸有一个以硫酯键连结的硫原子。
芳香族侧链	苯丙氨酸、酪氨酸和色氨酸带有可吸收紫外光的芳香族侧链。
相关主题	蛋白质结构与功能（B2）　　蛋白质合成机制（Q2）

结构

蛋白质是L-氨基酸的聚合物。除了脯氨酸外，蛋白质中已发现的所有氨基酸都有一个共同的结构，即与一个羧基相连的α-碳原子、一个氨基、一个质子和一个随氨基酸不同而异的**侧链**（R）（图B1.1）。除了甘氨酸外，所有的α-碳原子均具手性（不对称的），即它与四个不同的化学基团相连。因此氨基酸存在着成对的光学活性立体异构体（D-和L-）。然而在蛋白质中仅仅发现了L型异构体。在水溶液中氨基酸是**两性离子**，可以呈酸性和碱性（**兼性的**）。侧链可在大小、形状、电荷和化学活性上不同，这导致了不同蛋白质的性质各异（图B1.2）。一些蛋白含稀有氨基酸，如胶原中的4-羟基脯氨酸和5-羟基赖氨酸，它们是经脯氨酸和赖氨酸的**翻译后修饰**而形成的（参见Q4）。



图B1.1 L-氨基酸的一般结构。R基团表示侧链

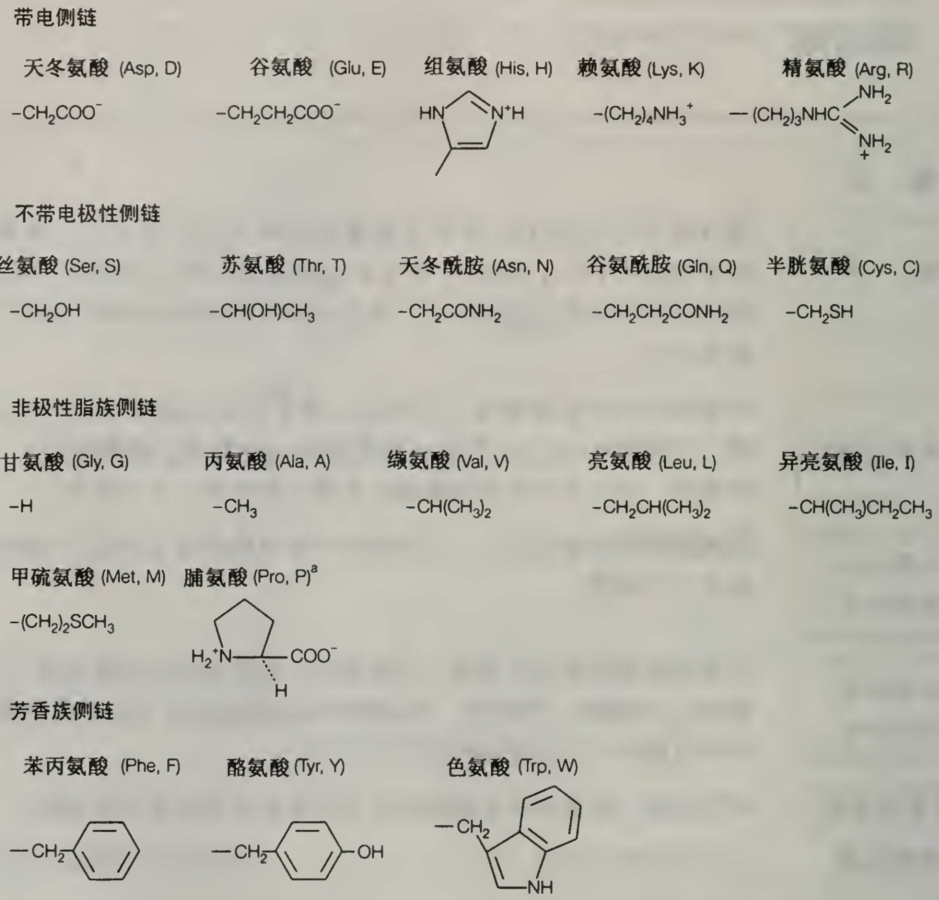


图 B1.2 20 种常见氨基酸侧链（R）。括号给出了标准三字母缩写和一字母缩写。a：脯氨酸结构显示出它是一个亚氨基酸

带电荷的侧链

以 pH 等于 7 为一个参考点，在此 pH 下几种氨基酸侧链上有提供额外的正或负电荷离子化的基团。酸性氨基酸**天冬氨酸**和**谷氨酸**，带有通常离子化的额外的羧基（带负电荷），而碱性氨基酸有带正电的基团，如**赖氨酸**有**亚氨基**连接在 ε-碳原子上，**精氨酸**则有一个胍基。**组氨酸**的咪唑基团的 pK_a 接近中性，在生理条件下这个基团可逆的质子化贡献于许多酶的催化机制的实现。酸性和碱性氨基酸可以组成蛋白质中重要的盐桥（参见 A4）。

不带电的极性侧链

不带电的极性侧链氨基酸含有可与水形成氢键的基团。与带电氨基酸一起，常被称为**亲水氨基酸**。**丝氨酸**和**苏氨酸**有羟基，而**天冬酰胺**和**谷氨酰胺**是由天冬氨酸和谷氨酸的胺衍生物。**半胱氨酸**有一个易被氧化硫醇基（巯基），巯基氧化后生成**胱氨酸**，其两个半胱氨酸的巯基间可形成在结构上很重要的二硫键（参见 B2）。

非极性脂肪烃侧链

甘氨酸的侧链位置有一个氢原子，不具光学活性。**脯氨酸**是一个亚氨基酸。

丙氨酸、**缬氨酸**、**亮氨酸**和**异亮氨酸**的侧链上带有**疏水性**烷基，在蛋白质结构中参与疏水相互作用（参见 A4）。**甲硫氨酸**在烷基侧链中有一个以硫酯键相连的硫原子。

芳香族侧链

苯丙氨酸、**酪氨酸**和**色氨酸**有较大的疏水侧链，其芳香结构承担蛋白质的紫外吸收，吸收在 280nm 最大。**酪氨酸**的酚羟基也能形成氢键。

B2 蛋白质结构与功能

要 点

大小与形状	包括许多酶在内的球蛋白，在溶液中呈紧密的、几乎球形的颗粒。纤维蛋白有一个高轴比，在结构形成上很重要，例如丝蛋白（fibroin）和角蛋白（keratin），大小范围可以从几千到几百万道尔顿。一些蛋白质常与非蛋白性物质相结合，例如脂类、糖或某些较小的辅因子。
一级结构	氨基酸在 α-羧基与 α-氨基间以肽键相连，形成的多肽链有一个 N 端和一个 C 端。以这种方式相连的多肽通常有 100～1500 个氨基酸。
二级结构	多肽可以折叠成几种规则结构，右手 α-螺旋每圈有 3.6 个氨基酸，在肽链的 N—H 与相隔三个残基的 C═O 基团间形成氢键以稳定其整体结构。平行和反平行的 β-折叠片层在肽链的不同部位间由氢键来稳定。
三级结构	不同的二级结构区域和连接区的组合折叠成一个确定的三级结构，结果是绝大多数亲水性氨基酸在外表面，而疏水性氨基酸在内部，其结构经非共价相互作用，有时是二硫键来稳定。变性往往导致二级和三级结构的丧失。
四级结构	许多蛋白质有多于一个的肽链亚基，如血红球蛋白有两条 α 链和两条 β 链。像微管这样的巨大复合体是由单个肽链的四元结合体组成的。别构效应通常取决于亚基的相互作用。
辅　　基	连接蛋白可与提供额外化学功能的非蛋白分子相结合。辅基包括烟酰胺二核苷酸（NAD^+）、血红素和金属离子（例如 Zn^{2+}）。
蛋白质的功能	蛋白质具有广泛的功能： • 酶可催化绝大多数生化反应，底物的结合取决于特异的非共价相互作用。 • 当与配体结合时，膜受体蛋白将信号传入细胞内部。 • 转运与贮存，例如血红蛋白在血液中转运氧、铁蛋白在肝脏内贮存铁。 • 胶原和角蛋白是重要的结构蛋白，肌动蛋白和肌球蛋白构成有收缩力的肌肉纤维。 • 酪蛋白和卵清蛋白是为生长提供氨基酸的营养蛋白。 • 免疫系统依赖抗体蛋白抵御感染。

• 调节蛋白诸如转录因子具有结合并调节其他分子（如 DNA）的功能。

结构域、基序、家族与进化

结构域可在一条肽链中形成半独立的结构和功能单位。结构域常由一个基因中的单个外显子编码。新的蛋白质可能是从外显子的新组合即由此产生的蛋白质的结构域的重新组合进化而来。基序（motif）是二级结构元件组合或在蛋白质家族的相关成员中常发现的氨基酸序列。相似的结构基序常在没有序列相似性的蛋白质中发现。蛋白质家族通过基因复制和随后的新基因趋异进化产生。

相关主题

生物大分子（A3）	氨基酸（B1）
大分子的组装（A4）	蛋白质合成（Q）

大小与形状

蛋白质可以区分为两大类球蛋白和纤维蛋白。**球蛋白**紧密折叠，在溶液中呈近似球状颗粒，自然界中的绝大部分酶是球形蛋白。**纤维蛋白**呈高轴比（长/宽），是重要的结构蛋白，例如头发与羊毛中的丝蛋白和角蛋白。蛋白质的分子质量可以从几千道尔顿（如胰岛素有 51 个氨基酸，约 5734Da）到几百万道尔顿，如丙酮酸脱氢酶这样的复合体至少有 500 万道尔顿。一些蛋白含有非蛋白结合物，可以小的**辅基**形式在酶反应中充当辅因子，或者是较大分子的联合（例如**脂蛋白**中的脂类或**糖蛋白**中的糖类，参见 A3）。

一级结构

氨基酸的 α-羧基通过酰胺键与下一个氨基酸的 α-氨基共价结合，在蛋白质中通常称为**肽键**。当两个氨基酸残基以这种方式相连，产物即称为**二肽**。许多氨基酸通过肽键相连形成**多肽**（图 B2.1）。α-碳原子和肽键的重复序列构成了多肽的骨架，而不同的氨基酸**侧链**授与蛋白质的功能。在多肽的一端的氨基酸有一个未结合的 α-氨基，而另一端有一个自由的 α-羧基。因此，多肽是有方向性的，有一个 **N 端**和一个 **C 端**。有时 N 端被像乙酰基这样的基团**封闭**。从 N 端到 C 端的氨基酸顺序就是多肽的**一级结构**。典型的单个多肽链的大小在 100～1500 个氨基酸范围之内，尽管也

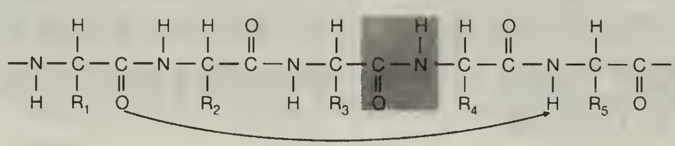


图 B2.1 一条多肽链的部分一级结构图。方框中图示肽键，在 α-螺旋中，第 *n* 个氨基酸残基的 CO 基团与第 *n* + 3 个氨基酸残基上的 NH 基团以氢键相连（箭头所示）

有更长或更短的多肽。

二级结构

肽键中 C═O 与 N—H 基团的高极性性质使 C—N 键具有部分双键特性，这使得肽键单元有刚性和平面性，尽管相邻肽键间存在自由旋转。这种极性有利于在适当的空间和方向上的肽键单元间形成氢键，因此多肽链能折叠成几种由氢键维持的规则结构。最熟悉的**二级结构**是 **α-螺旋**（图 B2.2a）。肽链骨架形成一个每周 3.6 个氨基酸的右手螺旋，每个肽的 N—H 基团都与相距 3 个残基的 C═O 基团间形成氢键（图 B2.1）。α-螺旋二级结构片段常在球蛋白和一些纤维蛋白中发现。**β-折叠片层**是由肽键 N—H 和 C═O 基团与肽链上另一段互补的基团间以氢键维系的（图 B2.2b）。多肽链中的几个片段可以肩并肩排列，形成一个带侧链（R）的片层结构，侧链突出于片层的上面或下面。如果这些片段走向相同（例如 N 端→C 端），片层是**正向平行**的；如果它们选择 N→C 和 C→N，那么片层是**反向平行**的。β-折叠坚硬牢固，是重要的结构蛋白，如丝蛋白。结缔组织蛋白——**胶原蛋白**有一个独特的**三股螺旋**的二级结构，其中三条肽链互相缠绕，使其十分牢固。

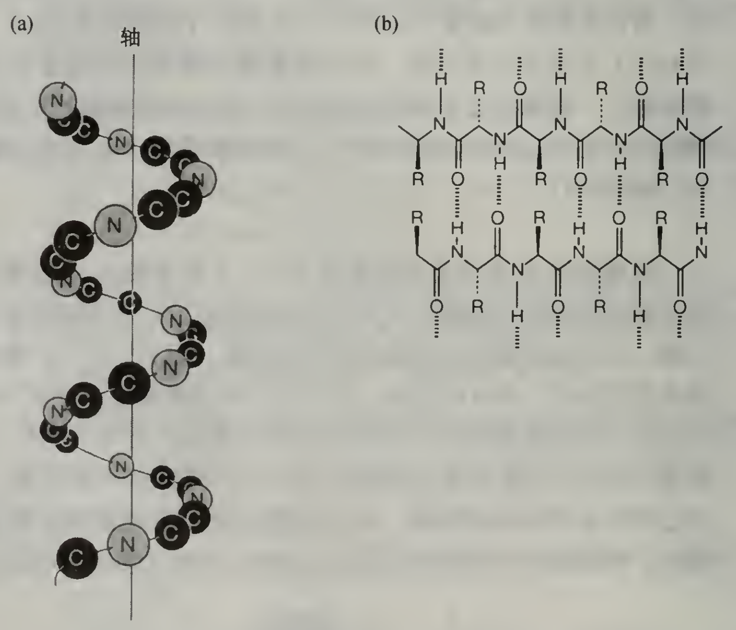


图 B2.2 二级结构。(a) α-螺旋，图中仅标出肽链骨架的 α-碳、肽键上 C 和 N 原子；(b) β-折叠

三级结构

不同片段 α-螺旋、β-折叠、其他微二级结构和连接环进一步折叠成三维构象的组织水平即多肽的**三级结构**（图 B2.3）。三级结构的特性存在于一级结构之中，给予合适的条件，绝大多数多肽将自发折叠成正确的三级结构，因为那通常是序列能量最低的构象。然而在体内正确的折叠，常

常需要称为**伴娘蛋白**（chaperone）的蛋白质帮助，伴娘蛋白在蛋白质合成（和一级结构）结束前阻止新生多肽的错误折叠。折叠就是：带有亲水侧链的氨基酸主要位于蛋白质外部，可与水或溶剂离子反应，而疏水氨基酸被埋在疏水的内部，这使得结构总体上稳定。侧链间各种类型非共价相互作用维系着三级结构，这些相互作用包括范德华力、氢键、带相反电荷基团之间静电盐桥（如赖氨酸 ε－NH_3^+ 基团和天冬氨酸或谷氨酸侧链 COO^-基团之间）以及存在于芳香族和脂肪族氨基酸非极性侧链间的疏水相互作用（参见 A4)。另外，共价二硫键可以在两个半胱氨酸残基间形成，这两个半胱氨酸在一级结构中可能相距很远，但在折叠的三级结构中会靠得很近。通过加热以及极端 pH 可以破坏二级和三级结构，这将导致蛋白质的**变性**和**无规卷曲**构象的形成。

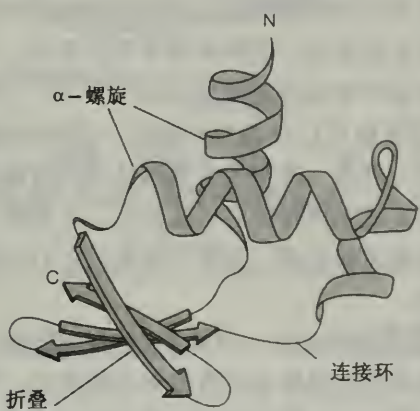


图 B2.3 蛋白质三级结构示意图

四级结构

许多蛋白质由两个或多个多肽链（**亚基**）构成。其多肽链可以是相同的或不同的。**血红蛋白**（hemoglobin）有两个 α 球蛋白链和两个 β 球蛋白链（$\alpha_2\beta_2$）。稳定三级结构的力同样可维系亚基在一起，包括不同多肽链的半胱氨酸间的二硫键。这一组织水平被称为**四级结构**，有着特定的重要性。首先，可构成很大的蛋白分子。微管蛋白是一个二聚体蛋白，由两个较小的不相同的 α 和 β 亚基组成。依靠微管结合 GTP 的水解，这些二聚体聚合成含有数以百计的 α 和 β 亚基的结构（参见 A4，图 A4.1），这些就是细胞骨架中的**微管**。其次，它能通过复合不同活性于一个整体赋予一个蛋白质以更多功能，如脂肪酸合成酶复合体。常常这些亚基间的相互作用通过与一些小分子的结合而被修饰，导致在酶调节中所看到的**别构效应**的发生。

辅基

许多**连接蛋白**可与被称为**辅基**的小分子以共价或非共价键相结合，给蛋白质提供氨基酸侧链所不能提供的功能。辅基的许多是酶催化反应中的**辅因子**，例如许多脱氢酶中的烟酰胺二核苷酸（NAD^+）、转氨酶中的磷

酸吡哆醛、血红蛋白和细胞色素中的血红素、金属离子（例如 Zn^{2+}）等。没有辅基的蛋白质被称为**脱辅基蛋白质**。

蛋白质的功能

- **酶**。除了少数有催化活性的 RNA 分子外（参见 O2），几乎所有的酶都是蛋白质。酶可以将生化反应速度提高几个数量级。**底物**结合涉及与特殊氨基酸侧链间的各种非共价相互作用。这些相互作用包括范德华力、氢键、盐桥和疏水力。结合的特异性极高，像只与单一底物结合（例如葡萄糖氧化酶只与葡萄糖结合），或者是基团特异性的（例如己糖激酶与各种己糖结合）。侧链也可以直接参与催化，例如可作为亲核体或质子供体或受体。
- **信号传递**。细胞膜上的受体蛋白质可以与来自细胞外介质的**配体**（如激素）结合，通过最终的构象改变，在细胞内启动对应于配体的反应。配体结合与底物结合很相似，但配体通常保持原形。一些激素本身是小分子蛋白质，如胰岛素和生长激素。
- **转运与贮存**。**血红蛋白**在红细胞中转运氧气，而**转铁蛋白**转运铁至肝脏，铁离子一旦进入肝脏，就与转铁蛋白结合以**铁蛋白**的形式储存于肝脏。摄入的脂肪在血液中以**脂蛋白**形式转运。许多其他分子和离子也都以蛋白质结合形式转运和贮存，这样在需要用它们之前，可以增加溶解性、减小反应活性。
- **结构与运动**。**胶原蛋白**是皮肤、骨骼和结缔组织中主要的蛋白质，而头发主要由**角蛋白**组成。细胞内还有许多结构蛋白，例如**细胞骨架**中的许多蛋白质。主要肌肉蛋白**肌动蛋白**和**肌球蛋白**组成了滑行丝，成为肌肉收缩的基础。
- **营养**。**酪蛋白**和**卵清蛋白**分别是奶和蛋中主要的蛋白质，被用来为发育中的后代生长提供氨基酸。种子蛋白质也为萌发的植物胚提供营养。
- **免疫**。可以识别并结合细菌、病毒和其他外源物质（**抗原**）的**抗体**，也是蛋白质。
- **调节**。**转录因子**与 DNA 结合并调节其功能。许多其他蛋白质通过与其他分子的结合修饰其功能。

结构域、基序、家族与进化

许多蛋白质由结构上独立的单元或**结构域**组成，结构域是在同一多肽中有限的高度有序结构片段相连而成。这种连接像铰链一样使单个结构域成相互关联地移动，用有限的蛋白水解方法破坏这种连接常可将结构域相分离，分离开的单个结构域可像单个球蛋白一样行动。酶的活性位点有时就形成于环绕底物的两个结构域间的沟中，结构域也可以有如与常用分子 ATP 结合的特殊功能。当许多不同蛋白质需要这一功能时，常会在它们中发现相同的结构域结构。在真核生物中，结构域常由基因中的不连续部分称为**外显子**来编码（参见 O3）。因此人们推测在进化中，经基因组中编码结构域的外显子的复制与重排产生结合位点、催化位点和结构元件的新

组合，从而最终产生出新的多肽。这样新的功能蛋白质的进化速率可以大大加快。

结构基序（也称为**超二级结构**）是频繁出现在球蛋白中的二级结构元件群，它们常常在功能上很重要，代表着蛋白质家族从其同祖先进化过程中保留下来的保守的结合位点或催化位点的必要部分；换言之他们代表着在不相关蛋白质凑在一起实现结构-功能统一的最佳解决方案。一个普通的例子是 **βαβ 基序**，其中 β-折叠片层的两个连续平行链间的连接是一个 α-螺旋结构（图 B2.4）。两个重叠的 βαβ 基序（βαβαβ）形成了许多不相关蛋白质中的二核苷酸（如 NAD^+）结合位点，其他基序仅由少数保守的、功能上重要的氨基酸组成，而不是超二级结构。

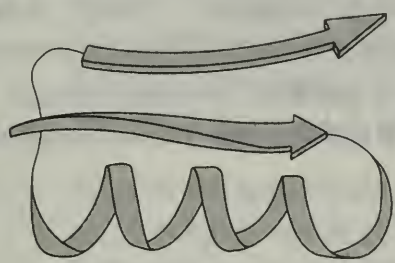


图 B2.4 βαβ 基序。α-螺旋用卷曲的带子表示，β-折叠片层用平箭头表示

蛋白质家族起源于一个古老基因的连续复制和随后的**趋异进化**。肌红蛋白（myoglobin）即肌肉中携氧蛋白、成人血红蛋白中 α-与 β-球蛋白链（和微小的 δ 链）以及胚性和胎儿血红蛋白中的 γ、ε 和 ξ-球蛋白都是**球蛋白家族**的相关多肽（图 B2.5）。这些蛋白和基因被称作是**同源的**（homologs）。不同物种的具有相同功能，承担相同生化角色的蛋白质家族成

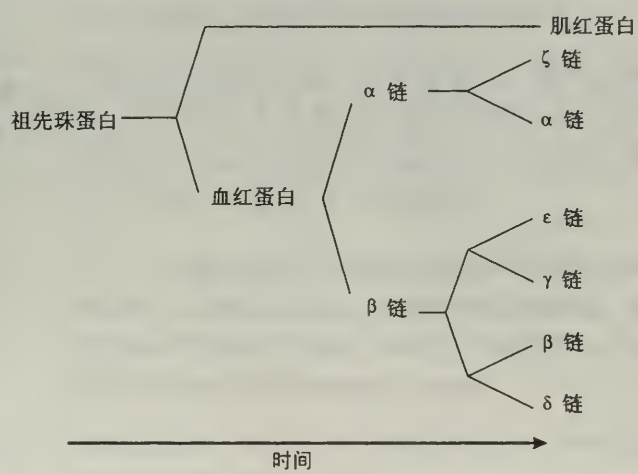


图 B2.5 来自同一个祖先珠蛋白基因的进化树

员（如大鼠和小鼠的肌红蛋白）为**直向同源**（定向进化同源 orthologs）。而那些进化不同但功能类似的蛋白（如 α-球蛋白与 β-球蛋白）为**共生同源**（平行进化同源 paralogs）。不同生物体的一个蛋白质家族直向同源成员间氨基酸序列的相似程度取决于两个生物体从共同祖先歧化的时间以及该序列的保守性对蛋白质功能的重要性。

一个蛋白质的功能，无论是催化或是结构性的（structural），均天生与结构相关。如上所述，由**趋同进化**（convergent evolution）可得到具有类似结构和功能的蛋白质，即无关基因进化至产生具有相似结构和催化活性的蛋白质。蛋白水解酶（细菌）与糜蛋白酶（功能）就是一个很好的例子。尽管它们的氨基酸序列差别很大，由不同的结构基序组合，但通过进化，他们氨基酸活性位点的**催化三联体**——丝氨酸、组氨酸、天冬氨酸——（catalytic triad）具有相同的空间取向使用几乎相同的催化机理去水解肽键。这样的蛋白质称为**功能相似**（functional analog），而相似结构基序独立进化，结果形成的**蛋白质结构相似**（structucral analog）。

B3 蛋白质分析法

要 点

蛋白质纯化	从细胞粗提物中纯化蛋白质可通过依据不同特性进行分离的方法组合来进行。凝胶过滤层析（gel filtration chromatography）是根据蛋白质的大小来分离；离子交换层析（ion-exchange chromatography）、等电聚焦（isoelectric focusing）和电泳均是根据蛋白质所带离子电荷的不同来分离；疏水相互作用层析利用了疏水性的不同；亲和层析依赖酶或者受体与配体间特殊亲和性，如底物与抑制物的亲和性来进行分离的。蛋白质的过量表达会大大提高产量，在重组蛋白中加入纯化标签（tag）可达到一步纯化。
蛋白质测序	用特异的蛋白酶或化学试剂将多肽切割成较小的肽段后，肽段可以在自动测序仪中利用 Edman 降解法（Edman degradation）从 N 端开始测序。通过不同特异性的蛋白酶切割产生的重叠片段的重排再现原始序列。为了得到一个蛋白质序列而对其基因或互补 DNA（cDNA）进行测序比对蛋白质测序要简单。
分子质量的确定	通过凝胶过滤层析和 SDS 存在下的凝胶电泳可以获取蛋白质的近似分子质量。利用电喷射电离（electrospray ionization，ESI）和基质辅助激光解吸电离（matrixassisted laser desorption/ionization techniques，MALDI）质谱可以给出小于 100kDa 蛋白质的准确分子质量。质谱也能检测翻译后的修饰。
X 射线晶体学与核磁共振	许多蛋白质可以结晶，利用 X 射线衍射可以确定它们的三级结构。溶液中小分子质量的蛋白质结构可以通过多维核磁共振（multidimensional nuclear magnetic resonance）来确定，尤其是正常的 ^{12}C 和 ^{14}N 被 ^{13}C 和 ^{15}N 取代后。
功能分析	对于一个蛋白质进行功能分析包括对它进行分离纯化，以及在突变体中由于基因突变或缺失而使蛋白质缺失，通过研究突变体的表现来研究蛋白质功能。有时也可通过将一个新蛋白的序列和结构与已知蛋白比较来预测它的功能。
蛋白质组学	在给定条件下对任何特定细胞类型的所有表达蛋白进行分析与鉴定即蛋白质组学。

相关主题 亚细胞器（A2） 蛋白质结构与功能（B2）

蛋白质纯化

一个典型的真核细胞可以含有数以千计的不同蛋白质，一些含量十分丰富，一些仅含有几个拷贝。为了研究某一个蛋白质，必须首先将该蛋白从其他蛋白质和非蛋白质分子中纯化出来。用于分离蛋白质的最重要特性有大小、电荷、疏水性和对其他分子的亲和性。通常采用多种方法的组合来实现蛋白质的完全纯化。

大小

凝胶过滤层析采用填充有特定孔径的颗粒悬浮液的柱子（分子筛）进行，将样品上样在柱子顶端时，比孔径大的蛋白质分子很快被洗脱至底部，因为它们被排除在颗粒之外，因而所需缓冲液洗脱体积相对较少。比孔径小的分子可以进入颗粒，因而需要相对较多的缓冲液洗脱体积。将小分子在大分子之后从柱上洗脱下来。超速离心时蛋白质在高速离心场中的沉降速度（**S 值**）随它的大小和形状不同而异（参见 A4）。超速离心不仅可以用于分离蛋白质而且也是研究蛋白质结构的有力分析工具。

电荷

由于蛋白质表面离子化侧链的存在，蛋白质带净电荷。由于这些侧链都是可滴定的（titratable），对于每个蛋白都存在一个 pH 使它的表面净电荷为零即**等电点**（pI）。在等电点时蛋白质最难溶。在等电点外的所有其他 pH 值，依据蛋白质所带净电荷采用**电泳**和**离子交换层析**来分离和纯化该蛋白质。在电泳中，蛋白质混合物被置于支持介质上，通常是一块凝胶，一个电场加于胶上，根据蛋白质所带的净电荷的多少将以不同的速度向阴极或阳极运动，分离后可从胶上回收。在离子交换层析中，结合在堆积于柱中的不溶的支持物上（离子交换剂）的带电离子，可以可逆地被溶液中带电蛋白质置换。不同的蛋白质对离子交换剂的亲和力不同，因而需要不同的离子强度再取代他们。通常用一个离子强度不断增长的盐浓度梯度通过柱子，结合的蛋白质可分别洗脱下来。在**等电聚焦**中，被称为**多聚两性电解质**的缓冲液混合物通过电场作用可以形成一个 pH 梯度。蛋白质在这个梯度中迁移至相应的等电点位置。随着净电荷趋于零运动停止而形成致密的聚集带（图 B3.1）。

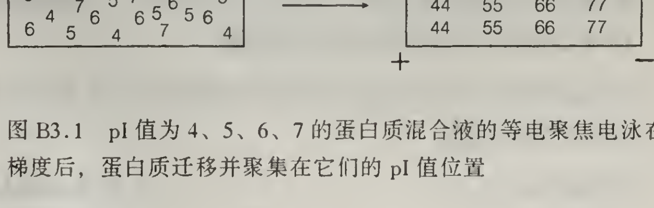


图 B3.1 pI 值为 4、5、6、7 的蛋白质混合液的等电聚焦电泳在电场中形成 pH 梯度后，蛋白质迁移并聚集在它们的 pI 值位置

疏水性

在高离子强度溶液中，蛋白质和柱中含有的芳香族或脂族烷基物质之间的疏水相互作用会增强。通过一个逐渐增加的离子强度梯度，蛋白质会在不同阶段被洗脱下来。

亲和性

酶与底物、受体与配体和抗体与抗原反应的高度特异性意味着可依据这些特性对蛋白质进行完全纯化。例如一个用与激素胰岛素共价结合的支持物制成的层析柱可以特异地结合胰岛素受体而不是别的蛋白质。胰岛素受体是一种低丰度的细胞表面蛋白质，可将胰岛素的生物活性传入细胞内部。由于酶的竞争性抑制剂不会被要被纯化的酶所降解，故常被用作亲和配体。

重组技术

在一个适合的宿主细胞中过量表达重组蛋白质，产量的提高大大简化纯化过程（参见 H1 和 J6）。例如将编码蛋白质的基因插入一个在强启动子控制下的噬菌体或质粒的**表达载体**（参见 K3），并将其转化大肠杆菌，该蛋白质的合成量可以占细胞总蛋白质的 30%。通过在基因的 5′或 3′端加入 6 个组氨酸密码子，形成用于纯化的标签，可进一步加速纯化进程。这种情况下被合成的蛋白质会在 N 或 C 端有一个六组氨酸标签，这使得在含有固定金属离子如可与组氨酸结合的 Ni^{2+} 的亲和柱上，对蛋白质进行一步纯化成为可能。标签通常对蛋白质功能没有什么影响。真核生物蛋白质可能需要在真核生物宿主-载体系统中表达以获得正确的翻译后修饰（参见 H4）。

蛋白质测序

理解蛋白质是如何行使功能的关键是了解其一级结构。蛋白质的氨基酸组成分析可以用酸（6mol/L HCl，110℃，24h）水解肽链，并经层析分离所得氨基酸的方法完成。这一方法仅给出了氨基酸的个数例如甘氨酸有多少、丝氨酸多少，并没有给出实际序列。序列的确定涉及要用特殊的蛋白水解酶或切割特定肽键的化学试剂将蛋白质分割成一系列的小肽（图 B3.2）。例如胃蛋白酶只切割赖氨酸（K）或精氨酸（R）之后的肽键，V8 蛋白酶只在谷氨酸（E）后切割；而溴化氰在甲硫氨酸残基后切割肽键。

然后将每一段小肽用自动蛋白质测序仪采用 **Edman 降解法**（Edman degradation）进行自动序列测定。其原理是异硫氰酸苯酯（phenylisothiocyanate，PITC）与 N 端氨基酸反应，在酸处理后释放乙内酰苯硫脲（phenylthiohydantoin，PTH）衍生物，产生新的 N 端。PTH-氨基酸经层析分离并与标准氨基酸相比而同定，重复循环可确定下一个氨基酸。原初

样品蛋白质的肽链顺序可以通过测定经不同蛋白酶处理产生出的肽段的序列，并根据重叠序列的办法推测出来（图 B3.2）。这种方法既费力又昂贵，因此大多数蛋白质的序列现在都是通过测定其基因或互补 DNA（cDNA）的 DNA 序列再根据遗传密码子来间接测定的（参见 J2）。此法比较简单且迅速，但不能反映翻译后修饰的状态（参见 Q4）。现在直接测序通常用于验证 N 端序列或某些限定的内部序列，这样的内部序列为构建用于寻找基因或 cDNA 的寡聚核苷酸或抗体**探针**提供信息（参见部分 I）。

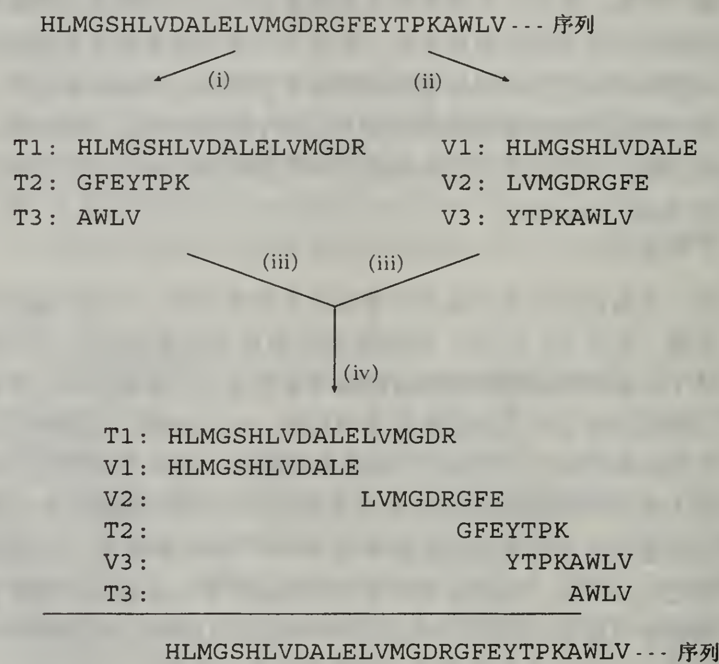


图 B3.2　多肽序列测定图解。(i) 胰蛋白酶裂解多肽链；(ii) V8 蛋白酶的裂解；(iii) Edman 降解法测定胰蛋白酶裂解的肽片段（T1～T3）和 V8 蛋白酶裂解的肽片段（V1～V3）的氨基酸序列；(iv) 根据重叠肽段组装完整的全长序列肽链

分子质量的确定

经凝胶过滤层析将样品蛋白质洗脱时间与已知标准蛋白的洗脱时间相比可给出样品蛋白的近似分子质量。SDS 聚丙烯酰胺凝胶电泳（SDS-PAGE）也可以用于确定单个肽链的大小（在离子型去污剂 SDS 的存在下，肽键失去其四级结构），因为 SDS 离子型硫酸盐赋予所有蛋白质依赖于质量的负电荷，此时电场中蛋白质的移动速度取决于其质量而不是电荷。这些方法尽管不够精确（误差 5%～10%），但很方便且成本低。**质谱**可是一个极为准确的测定方法。分子被 Xe 或 Ar 离子束气化、离子化，电磁场中离子折射程度是依赖其质量的并可以被测量。然而这一方法的测定上限仅几 kDa，而对蛋白质是破坏性的。在这一数量级范围内，现在非破坏性离子化技术得到了极大发展。**电喷射电离**（SEI）法产生一个高度带电的蛋白质溶液滴的细雾，气化形成带电气态蛋白质离子；而**基质辅助**

激光解吸电离（MALDI）质谱是通过激光气化含有蛋白质的固态基质而生成气相离子。对于小于100kDa的蛋白质，用质谱来测量其分子质量精确度可达0.01%。质谱仪可用于测翻译后修饰，最近也开发出有助于蛋白质鉴定以及序列分析的技术。

X射线晶体学和核磁共振

球形蛋白质有确定的三维（3-D）结构，因此许多球形蛋白质可以结晶。可用**X射线晶体衍射**来确定蛋白质的三维结构。X射线与蛋白质中的电子作用，当一束射线通过晶体时通过测量X射线的衍射图像，可以计算出晶体中原子的位置。在底物存在下结晶一种酶，就可以看到结合和催化反应中精确分子间的相互作用。溶液中小分子球蛋白结构也可以用二维或三维**核磁共振谱**来确定。核磁共振中质子在强磁场中被射频激发后产生的质子松弛谱（relaxation）可被测定。松弛谱的特性取决于分子中质子的相对位置。用于蛋白质结构分析的多维NMR需要进一步发展并解决好大量质子产生的叠加数据。用^{13}C和^{15}N取代蛋白质中正常的^{12}C与^{14}N可消除不必要的共振可大大地改善数据解析。用这种方法大至30kDa的蛋白质结构都可以被测定。用X射线和核磁共振方法来确定蛋白质结构，两种方法的结果通常吻合得很好。这表明测定的结构就是蛋白质在体内的真实结构。

功能分析

许多蛋白的三维结构现在已经确定。这些结构信息对于新型药物的开发有很高价值，可用来设计与目标蛋白具有高度亲和性、可特异结合的新型药物。但是，确定三级与四级结构仍是一个相当花钱和困难的过程，且远远滞后于基因组测序计划所推测的新蛋白一级结构的需求（参见J2）。因此现在计算机方法引起了人们很大兴趣，应用计算机技术可以从简单的氨基酸序列信息预测蛋白质结构和可能的功能。这些方法很大程度上是基于以下条件：在自然界中仅存在有限数量的超二级结构，对新蛋白序列用相关氨基酸序列的已知三维结构的蛋白进行作图，不过在解释这样的结果时要特别小心。例如：溶菌酶与α-乳清蛋白显示出高度的序列同源（约70%）和结构相似性，明显有着密切关系。尽管α-乳清蛋白失去了溶菌酶中可水解碳水化合物的关键的催化氨基酸残基，结果仅能与之结合，就像根据氨基酸序列相似性预测的结果那样它并不是一个酶。因此，至少在现在，了解一个蛋白质的真正功能，仍需要对它进行分离，进而了解其生化与结构特征。基因重组技术在很大程度上有助于蛋白质分离。运用这些技术可以克隆表达基因组测序计划所确定的新基因，并制造出那些过去从未被检测过的一些蛋白质（参见J2）。功能分析的另一个重要方面是确定细胞中多个蛋白质与所有其他蛋白质的相互作用。

一个蛋白质的生化特征不会告诉你该蛋白质在细胞中起什么作用，往往需要更多的遗传分析。如果某一蛋白质的基因可通过突变失活或用重组DNA技术删除，进而研究产生的突变体的**表现型**（phenotype）。通过与

已知的生化信息一同考虑，突变细胞的不同表现有助于揭示该蛋白质在体内的功能。现在已经可以制作酿酒酵母的全部 6000 种蛋白质或编码这些蛋白质的基因的突变体，生产出的一套突变体将有助于确定这一简单真核生物中各蛋白质所起的作用。

蛋白质组学

与基因组（genome）一词相似，单词 **transcriptome** 用于描述由一个细胞的基因组转录的全部 mRNA（参见 K1），而**蛋白质组**（proteome）则用于描述一个细胞在它的生存期或它生存的任何一个时间内转录表达的全套蛋白。蛋白质组也包括单个基因由于转录后 RNA 的不同剪切（参见 O4）以及某一蛋白翻译后经修饰所产生的全部不同产物（参见 Q4）。虽然一个生物体内细胞间基因组是相当一致的，而转录组与蛋白质组在各类细胞之间（如脑与肝）差异很大，甚至单一细胞类型的不同细胞周期内（参见 E3）也是如此；或是由于暴露于不同的外界刺激（如激素、生长因子、压力）使基因表现型产生很大变化。

蛋白质组学就是用高分辨率的蛋白质分离和鉴定技术来研究蛋白质组。当前，最好的分离方法是**二维凝胶电泳**。从一个细胞或组织中所分离出的蛋白质首先在聚丙烯酰胺凝胶中由所带电荷在**等电聚焦**作用下形成窄带而得以分开，再将凝胶旋转 90°，蛋白质在一块含有 SDS 的凝胶电泳中，由于分子质量不同而再次得以分开。结果可得到一张可分开几千种蛋白系列的蛋白斑点的二维图谱。各个点迹被切下来，用蛋白水解酶如胰蛋白酶处理，得出该蛋白质的一套肽谱。样品中每一个特定肽段的精确分子质量可由 MALDI 质谱仪确定，产生出该蛋白质的一个**多肽分子质量指纹图谱**。现在已有许多生物体的丰富的 DNA 序列信息，并已建立起根据这些序列所推测的指纹图谱数据率，再将该蛋白质的指纹图谱与之比较。以上的诸多程序可用机器自动完成，一个给定细胞的蛋白质组的大量蛋白质均可判定出来。这样的蛋白质定位对于我们理解细胞如何工作以及在疾病中细胞功能的改变极为重要，并对新药设计药物化学有着重大意义。药物公司将会产生很大兴趣，以满足他们在新药设计时对蛋白质信息的要求。

（董晓楠 译 刘进元 校）

C1 核酸结构

要 点

碱 基	DNA 含有 4 种杂环碱基：嘌呤类有腺嘌呤（A）和鸟嘌呤（G），嘧啶类有胞嘧啶（C）和胸腺嘧啶（T）；而在 RNA 中尿嘧啶（U）替代了结构非常相似的胸腺嘧啶。
核 苷	核苷由碱基共价结合于戊糖分子的 1′位而构成。RNA 中的戊糖为核糖（ribose），构成核糖核苷（ribonucleoside）或简称核苷；而在 DNA 中戊糖为 2′-脱氧核糖（2′-deoxyribose），形成 2′-脱氧核糖核苷（2′-deoxyribonucleoside）或简称脱氧核苷。碱基 + 糖分子 = 核苷。
核苷酸	核苷酸由一个或多个磷酸基团共价结合到核苷的 3′、5′位，或 2′位（在核糖核苷中）而形成，即碱基 + 糖分子 + 磷酸分子 = 核苷酸。RNA 和 DNA 分别由相应的 5′-三磷酸核苷，即 5′-三磷酸核糖核苷（NTP）和 5′-三磷酸脱氧核糖核苷（dNTP）构成。
磷酸二酯键	核酸链中，核糖或脱氧核糖的 5′位和与其相邻的戊糖的 3′位之间由磷酸键连接，即形成 3′，5′-磷酸二酯键。因此核酸由具有方向性的糖-磷酸骨架，以及结合于每一糖分子 1′位的碱基构成。其重复单位是核苷酸。由于磷酸分子带有负电，致使核酸成为具有强负电性的多聚大分子。
DNA/RNA 序列	核酸序列是指 DNA 或 RNA 链中的碱基 A、C、G、T 或 U 排列顺序，习惯上由分子游离的 5′端写至 3′端，例如，5′-TATTGCTC-3′（DNA）或 5′-AUAGCUUGA-3′（RNA）。
DNA 双螺旋	DNA 分子通常以双螺旋形式存在。两条独立的反向平行的单链 DNA 分子以右手螺旋方式相互缠绕，糖-磷酸骨架在外，靠氢键和堆积力相互配对的碱基在内。腺嘌呤（A）与胸腺嘧啶（T）配对，鸟嘌呤（G）与胞嘧啶（C）配对。两条链是互补的，一条链决定了另一条链的序列。
A 型、B 型及 Z 型螺旋	虽然沃森与克里克所发现的标准的 DNA 双螺旋，即 B 型被认为是细胞中 DNA 占优势的结构，核酸也可以形成右手 A 型螺旋，这是 RNA 链所采用的形式；而左旋 Z 型只是在特殊碱基序列处形成，并在细胞中或许不是一个重要的构象。

RNA 二级结构

大部分 RNA 分子以可折叠成复杂构象的单链形式存在，这种复杂构象是由局部分子内碱基配对和其他氢键相互作用而维持的。这种复杂性反映出 RNA 在细胞中担负着多种功能。

修饰核酸

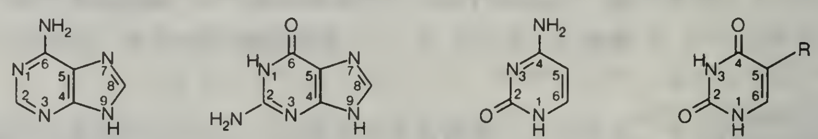
核酸的共价修饰在细胞中具有特定作用。这些修饰对 DNA 而言通常仅局限于腺嘌呤和胞嘧啶的甲基化（methylation）；而对 RNA 而言，修饰范围要大得多。

相关主题

核酸的理化特性（C2）
原核与真核生物的染色体结构（D）
核酸的光谱学和热力学特性（C3）
DNA 超螺旋（C4）

碱基

DNA 和 RNA 中的**碱基**是含有多种取代基的杂环（含 C、N）芳香族化合物（图 C1.1）。**嘌呤**类为双环结构，包括腺嘌呤和鸟嘌呤；而**嘧啶**类为单环结构，包括胞嘧啶、胸腺嘧啶和尿嘧啶。在 RNA 中尿嘧啶替代了胸腺嘧啶。胸腺嘧啶与尿嘧啶的不同仅在于其 5′位含一甲基，因此也可称胸腺嘧啶为 5-甲基尿嘧啶。



腺嘌呤 (A)　　鸟嘌呤 (G)　　胞嘧啶(C)　　尿嘧啶 (U; R = H) 胸腺嘧啶 (T; R = CH_3)

图 C1.1　组成核酸的碱基

核苷

核酸分子中，碱基共价结合于戊糖的 1′位构成**核苷**（图 C1.2）。RNA 中的戊糖为**核糖**，而在 DNA 中为 **2′-脱氧核糖**，即 2′位的羟基为氢原子所替代。核糖与碱基的结合位置，嘧啶为 1 位（*N* - 1），嘌呤为 9 位（*N* - 9）（图 1）。为了与碱基序位相区别，戊糖环中的原子数标作 1′-、2′-等。碱基与糖分子之间的结合键称为**糖苷键**（glycosylic or glycosidic bond）。若糖分子是核糖，形成的核苷（术语上应称为核糖核苷）为腺苷（adenosine）、鸟苷（guanosine）、胞苷（cytidine）和尿苷（uridine）；若糖分子是脱氧核糖（在 DNA 中），则形成的核苷（2′-脱氧核糖核苷）称为脱氧腺苷（deoxyadenosine）等等。胸苷（thymidine）和脱氧胸苷（deoxythymidine）二词，可以互换使用。

核苷酸

核苷酸由一个或多个磷酸分子共价结合于核苷的核糖的 3′、5′位或 2′位（仅在核糖核苷中）而成。若糖分子是脱氧核糖，则形成的核苷酸称为**脱氧核糖核苷酸**（deoxynucleotide）（图 C1.3），属于磷酸酯（phosphate

esters）类化合物。在5′位上，最多可以连接3个磷酸基团，像5′-三磷酸腺苷（adenosine 5′-triphosphate）或5′-三磷酸脱氧鸟苷（deoxyguanosine 5′-triphosphate），通常可简写为ATP和dGTP，同样还有dCTP、UTP和dTTP等。5′-单磷酸核苷和5′-二磷酸核苷可简写成比如AMP和dGDP等。5′-三磷酸核苷（NTP）或5′-三磷酸脱氧核苷（dNTP）是组成核酸大分子的基本元件。在DNA或RNA合成过程中，每个核苷酸脱掉一个焦磷酸基团（pyrophosphate）（含两个磷酸根），而只保留一个磷酸基团，以此整合到核酸链中（参见E1和K1）。因此DNA和RNA链中的重复单位是单核苷酸。

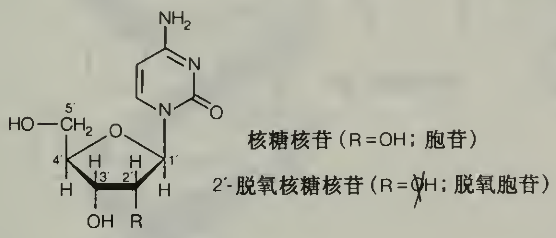


图 C1.2 核苷

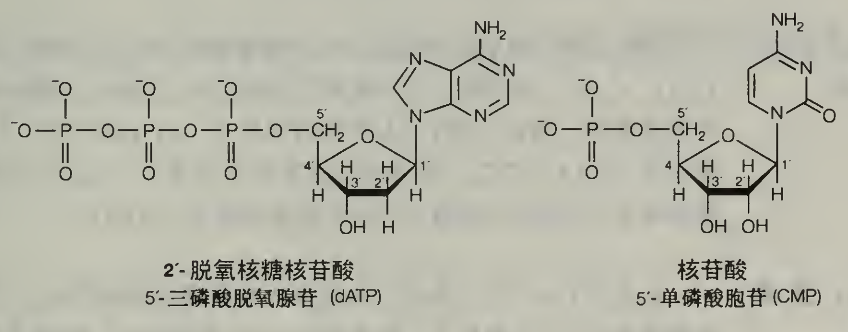


图 C1.3 核苷酸

磷酸二酯键

在DNA或RNA分子中，脱氧核糖核苷酸或核糖核苷酸通过一个磷酸基团与前一核糖的5′-羟基和下一个核糖的3′-羟基的共价连接而形成多聚物（图C1.4）。这种键或者连接称为**磷酸二酯键**，因为磷酸在化学上成了双酯形式。因此核酸链可被认为具有方向性。任何一条核酸链（环状的除外，参见C4），无论其长度如何，都具有一个连接或没有连接着磷酸基团的游离5′端（free 5′- end），和一个通常含有游离羟基的游离3′端（free 3′- end）。在中性pH条件下，每一磷酸基团带有一个单位负电荷。这就是核酸被称为“酸”的原因，是强酸的阴离子，所以核酸是带有**强负电性的多聚体**。

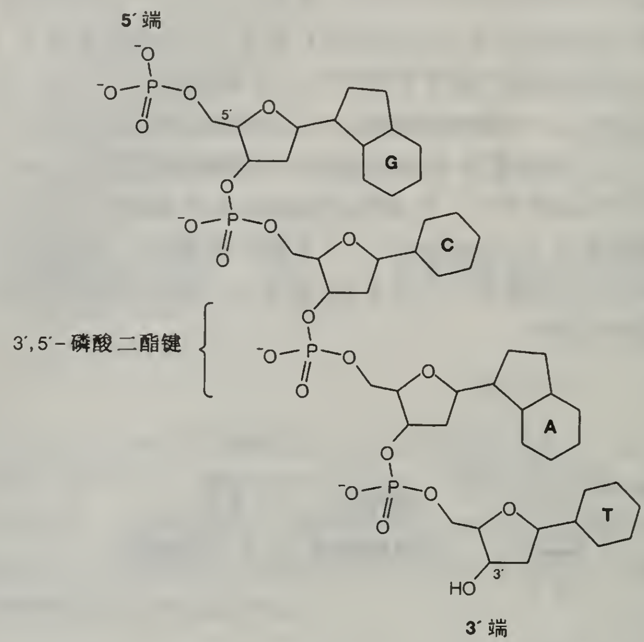


图 C1.4 磷酸二酯键与 DNA 链的共价结构

DNA/RNA 序列

习惯上将 DNA 或 RNA 分子的重复单体用它们的单个碱基字母 A, T, G, C 或 U 来代表。还有通常从左向右按 5′端向 3′端的方向书写核酸的碱基顺序。例如一段 DNA 序列可写成 5′-ATAAGCTC-3′,或者甚至直接写成 ATAAGCTC,一段 RNA 的序列可写成 5′-AUAGCUUGA-3′。注意链是有方向性的,例如 ATAAG 并不等同于 GAATA。

DNA 双螺旋

自然界 DNA 分子通常以众所周知的**双螺旋**形式存在。这种结构的基本特征是由 J. 沃森和 F. 克里克于 1953 年提出的,即两条相互独立的单链 DNA 分子以螺旋方式相互缠绕,形成**右手**双螺旋(图 C1.5a)。带有负电的戊糖-磷酸骨架在分子的外侧,碱基则平面堆积于螺旋的内部(图 C1.5b)。由于骨架双链在螺旋轴上的间距并不相等,从而在分子表面形成宽窄不等的**大沟**(major groove)和**小沟**(minor groove)。双链间对应的碱基靠氢键相互作用形成**碱基对**(base pair)。DNA 双螺旋结构中每一匝螺旋含约 10 个碱基对。两条链以 5′→3′方向,反向排列(反向平行),更重要的是在序列上两条链互为**互补链**(complementary)。这样的特征是由于碱基结构和 DNA 骨架限制支配着,碱基间以氢键形成嘌呤-嘧啶配对,因为它们间具有十分相似的空间几何构型(图 C1.6),G 和 C 配对(3 个氢键),A 和 T 配对(2 个氢键)。这样任意的序列均可构成规则的 DNA 双链结构。双链中一条链的序列可特异地确定另一条链的序列,这些均暗示出 DNA 自我复制(replication)和 DNA 转录(transcription)为

RNA 的机制（参见 E1 和 K1）。

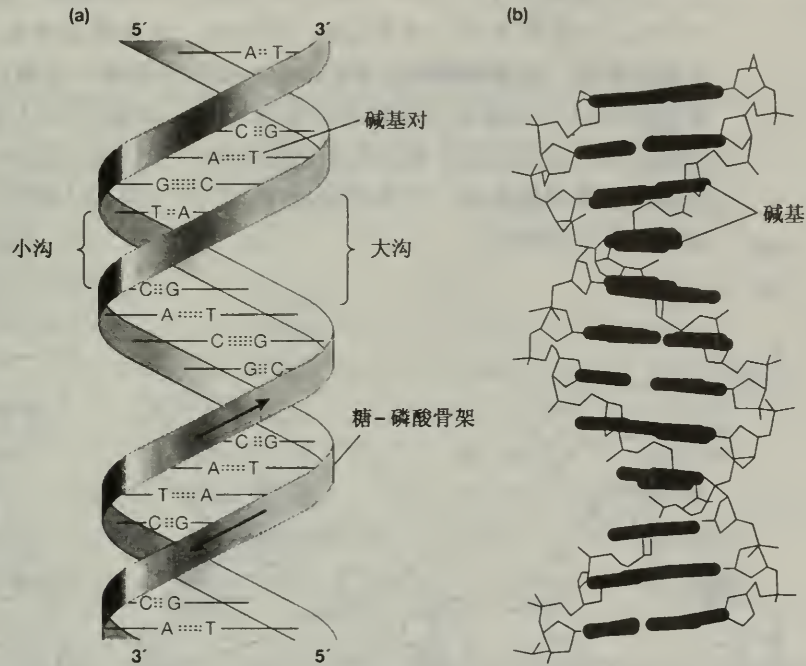


图 C1.5 DNA 双螺旋结构。(a) 结构图；(b) 结构细图（粗线表示碱基对的堆积）

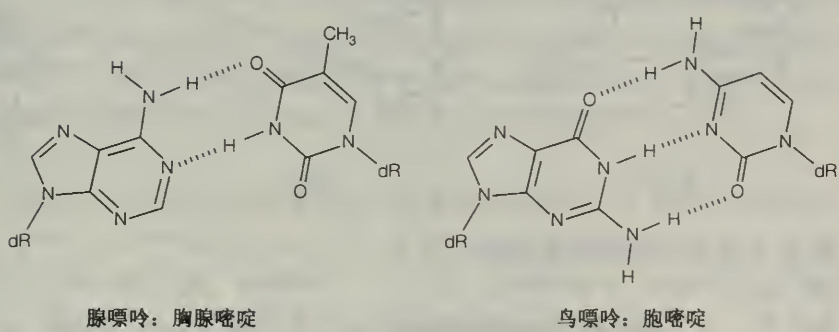


图 C1.6 DNA 碱基对。氢键用虚线表示；dR＝脱氧核糖

A 型、B 型及 Z 型螺旋

事实上，人们观察到有一些不同形式的核苷酸双链螺旋，并进行了研究，这些螺旋均有基本的反向平行的两条螺旋链的形式。该结构由沃森与克里克确定，如上所述，即 **B-DNA**（图 7a），被认为是适用于所有 DNA 的稳定形式。它的特殊的螺旋重复（helical repeat）每匝为 10bp 碱基对（尽管现在认为真正的 DNA 螺旋重复近于每匝 10.5 碱基对）。现已确认碱基对分布于螺旋轴线上，几乎与它垂直。B-DNA 有一条主沟和一条小

沟。

在低温条件下，DNA 可被诱导形成另一种螺旋，称为 **A 型**（图 C1.7b）。A 型与 B 型一样为右旋，但较宽，结构更加紧密。碱基对倾斜于螺旋轴线，且偏离细线（见文末，A 型螺旋中有一空洞）。A 型螺旋重复每匝为 11 个碱基对。尽管在异常环境中，细胞中 DNA 会形成 A 型或近似 A 型，A 型的主要意义在于它是 RNA（见下文）及 RNA 与 DNA 杂合体的主要螺旋形式，并说明将 RNA 的 2′-OH 纳入理论上更稳定的 B 型结构是不可能的。

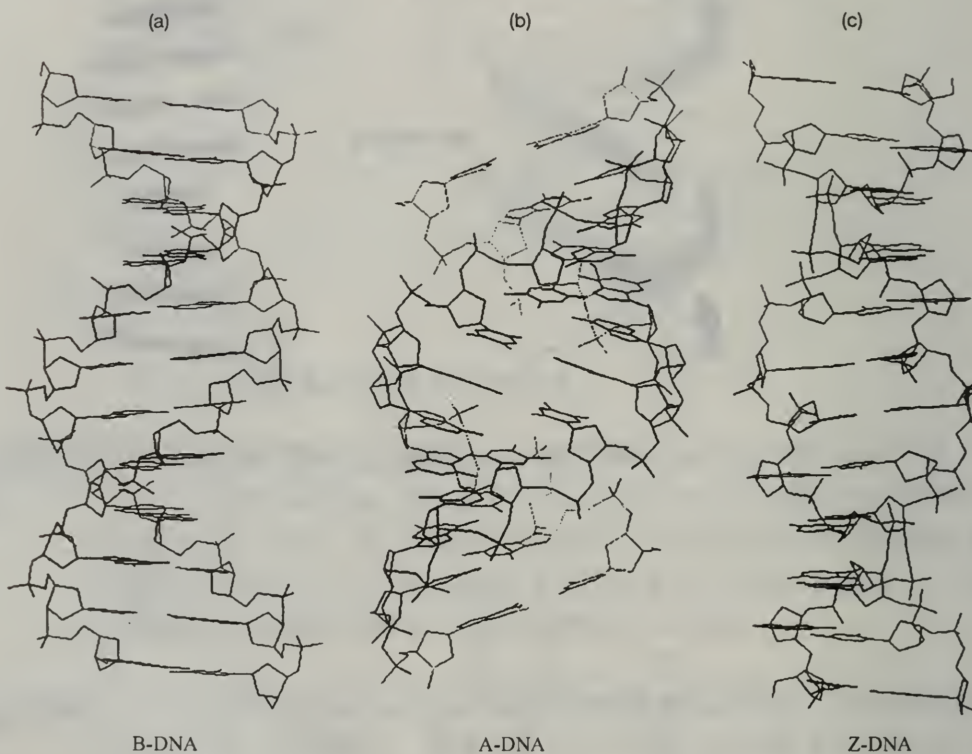


图 C1.7　DNA 双螺旋的不同螺旋模式

另一种 DNA 的异常螺旋形式是左旋 Z-DNA。在单一交替的嘧啶-嘌呤序列（如 5′-CGCGCG-3′，当然另一条链也如此）的合成双链 DNA 中，形成的左旋 **Z-DNA**（图 C1.7c）是稳定的。这是因为在这种结构中，嘧啶与嘌呤形成差别很大的构象，而不像在 A 型与 B 型，每个核苷酸均有相同的构象和瞬时环境，特别在 Z 型螺旋中，嘌呤核苷酸形成 *syn* 构象，嘌呤核苷酸及 A、B 型中所有核苷酸均为 *anti* 构象。Z 型螺旋有一个 Z 字型的外观，每匝 12 个碱基对。尽管它给人的感觉是每匝为 6 个碱基对的二聚体，但沿每条链的重复单位的确是一个二核苷酸。因为左旋 Z 型与 B 型之间的界限非常不稳定，即使是在重复的 CGCGCG 区域，Z-

DNA 也不易形成正常型的 DNA。尽管很有趣，Z 型不是体内 DNA（或 RNA）的主要形式（表 C1.1）。

表 C1.1 A、B 及 Z 型螺旋的主要特征概括

	A 型	B 型	Z 型
螺旋方向	右手	右手	左手
直径	约 2.6nm	约 2.0nm	约 1.8nm
每匝螺旋碱基对数目（*n*）	11	10	12（6 个二聚体）
每对碱基旋转角度（=360/*n*）	33°	36°	60°（每个二聚体）
每对碱基旋转上升（*h*）	0.26nm	0.34nm	0.37nm
螺距	2.8nm	3.4nm	4.5nm
碱基对与中心轴倾角	20°	6°	7°
大沟	窄，深	宽，深	平坦
小沟	宽，浅	窄，深	窄，深
糖苷键	*anti*	*anti*	*anti*（嘧啶） *syn*（嘌呤）

RNA 二级结构

RNA 分子通常以单链形式存在，因此不具有双链 DNA 分子那样的规则双螺旋结构，而是相对来说形成近似于球状的构型，通过**分子内**的氢键作用和在单核酸链间的碱基堆积可形成单链局部区域的螺旋结构。在这些区域中 RNA 分子的一部分和与其配对的另一部分互补（参见 P2，图 C1.2）。与 DNA 相比，RNA 分子的这种构型的多样性是与其在细胞中的作用的多样性相适应的。RNA 的结构可包括从较短的 snRNA 到较大的 rRNA 分子，前者在真核细胞内有助于 mRNA 前体的拼接（参见 O3），而后者则形成核糖体的骨架结构，并参与蛋白质的化学合成（参见 Q2）。

修饰核酸

核酸分子中的碱基或核苷酸的化学修饰是普遍的，并起着许多特定作用。虽然在某些噬菌体 DNA 上发生着更加复杂的修饰，但对于细胞 DNA，修饰一般仅限于腺嘌呤的 N-6 位、胞嘧啶的 4-氨基和 5 位（图 C1.1）的甲基化。这些甲基化具有限制性修饰作用（参见 G3）、碱基错配修复作用（参见 F3）且在真核基因组构中普遍存在（参见 D3）。更为复杂的修饰发生在转录后的 RNA 分子中，这又一次反映了 RNA 在细胞中的不同功能。更为详细的叙述参见 O3 和 P2。

C2 核酸的理化特性

要 点

核酸的稳定性	由于氢键的作用，DNA 双链和 RNA 结构似乎是稳定的。然而事实并非如此，氢键决定了碱基间的特异配对，而核酸螺旋的稳定性则是疏水作用（hydrophobic interaction）和堆积在碱基对间的偶极矩作用（dipole-dipole interaction）的共同结果。
酸效应	强酸条件下核酸可水解为碱基、糖和磷酸；中度的酸性可使嘌呤的糖苷键水解而产生脱嘌呤核酸（apurinic acid）。更为复杂的化学过程已被发展用来移去特定碱基，这也是 DNA 测序的基础。
碱效应	强碱条件下 DNA 和 RNA 因其碱基的互变异构态（tautomeric state）的改变以及特异氢键被破坏而变性。强碱条件下，RNA 也易于发生水解，2′- OH 参与磷酸二酯键骨架的分子内裂解。
化学变性	某些化学试剂，如尿素和甲酰胺，在中性 pH 条件下能够破坏堆积于碱基间的疏水作用而使 DNA 和 RNA 变性。
黏性	DNA 分子很细长，因此其溶液具有较高的黏性。溶液中的长 DNA 分子易被剪切而断裂，这种过程可以被用来产生特定长度的 DNA 片段。
浮力密度	DNA 的密度约为 $1.7 g \cdot cm^{-3}$，可以通过氯化铯密度梯度离心法加以分析和纯化。DNA 分子的精确密度是其（G+C）含量的函数，这项技术可用来分析 DNA 的碱基组成。
相关主题	核酸结构（C1）　　核酸的光谱学和热力学特性（C3）

核酸的稳定性

乍一看，由于碱基间氢键的存在使得 DNA 双链和 RNA 二级结构趋于稳定，但事实并非如此，正如在蛋白质（参见 A4）结构中的氢键通常并不足以提供这种稳定性一样。这是因为，例如对于 DNA 分子必须考虑到其单链的无规则卷曲与双链构型之间的能量差别。双链分子中碱基与碱基的氢键作用仅仅替代了单链分子中碱基与水分子的氢键作用，这两种作用从能量的角度讲是相同的。当然，氢键使得双链中的碱基间的配对具有特异性（只有互补的两条链之间才能形成 DNA 双链，参见 C1），但其对于双螺旋的总体上的稳定性并无贡献。核酸分子的稳定性的根源在于碱基对之间的**堆积作用**（stacking interaction，参见图 C1.5b）。作为芳香族化合物，碱基的平面使其不能在自由溶液中与水分子形成氢键，换句话说它们是**疏水的**。大量水分子的氢键作用而形成的网络在疏水的表面附近变得

不稳定，并非所有的水分子都能参与氢键相互反应，而且这些水分子变得更为有序。因此将碱基对堆积起来，从而使所有的水分子排除在这样的表面之外，从能量的角度讲是最稳定的：更多的水分子结合于巨大的氢键网络中。同时这样的过程也最大限度地增强了碱基的**电荷偶极作用**（参见A4）。即使是在单链 DNA 中，碱基也存在着互相堆叠的趋势，但这种堆叠在双链 DNA 中达到最大化，**疏水效应**使其成为能量上最为稳定的一种结构。事实上，以上论述是将一种极为复杂的情况简单化了，更为深入的解释则超出了本书的范围。

酸效应

在强酸和高温，如在高氯酸（$HClO_4$）中 100℃ 以上的条件下，核酸完全**水解**为碱基、核糖或脱氧核糖和磷酸。在浓度略稀的无机酸中，如 pH 3～4，最易水解的化学键被选择性地断裂，一般为连接嘌呤和核糖的糖苷键，从而产生**脱嘌呤核酸**（参见 F2）。更为复杂的化学过程已被开发用来移去某些碱基，或将 DNA、RNA 在特定碱基处断裂。以其发明人 **Maxam 和 Gilbert** 的名字命名的 DNA 化学测序法就是以此为基础的（参见 J2）。

碱效应

DNA

当 pH 值超出生理范围（pH 7～8）时，对 DNA 的结构将产生更为微妙的影响，碱效应使碱基的**互变异构态**发生变化。这种效应可以化合物环己酮（cyclohexanone）（图 C2.1a）为例加以说明：该化合物分子在互变异构态酮式和醇式间达到平衡。中性 pH 条件下，该化合物主要以**酮式**存在；pH 升高时，由于失去一质子，导致分子向**烯醇式**转化，这是因为电负性氧原子的存在，使分子在带负电的情况下达到稳定。同样的，在高 pH 下，嘌呤的分子结构（图 C2.1b）转为烯醇式，而其他的碱基也有类似的变化。这种变化影响到特定碱基间的氢健作用，结果导致 DNA 双链

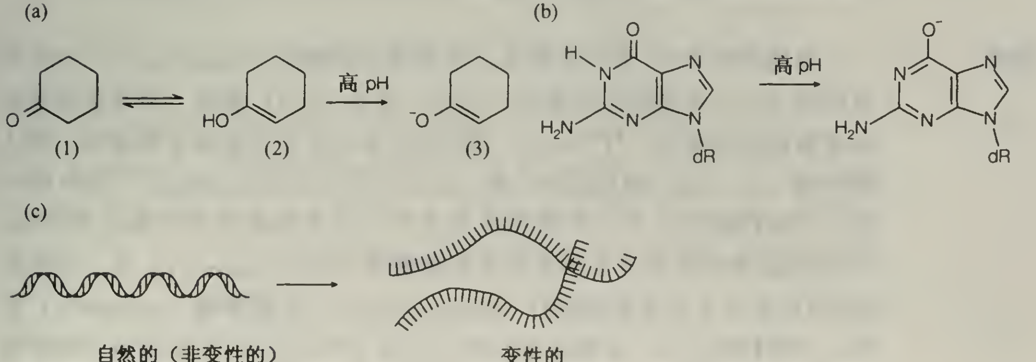


图 C2.1 高 pH 条件下 DNA 的变性。(a) 碱使之变成烯醇式异构体；(b) 脱氧嘌呤的异构体互变；(c) 双螺旋 DNA 的变性

解离，称为 DNA 的**变性**（图 C2.1c）。

RNA

pH 较高时，同样的变性发生在 RNA 的螺旋区域中，但通常被 RNA 的碱性**水解**所掩盖。这是因为 RNA 中存在的 2′-OH 参与到对磷酯键中磷酸分子的分子内攻击，从而导致 RNA 的断裂（图 C2.2）。因为氢氧根（-OH）是一个碱基，这一反应可以被高 pH 所驱动，反应产物是 5′-OH 和 **2′，3′-环磷酸二酯**（2′，3′-cyclic phosphodiester），后者继而分解为 2′-或 3′-单磷酸。即使在中性 pH 下，由于存在有 2′-OH，RNA 对水解的敏感程度也比 DNA 高得多。这就解释了为什么 DNA 分子中存在 2′-脱氧核糖，DNA 的功能要求其具有较高的稳定性。

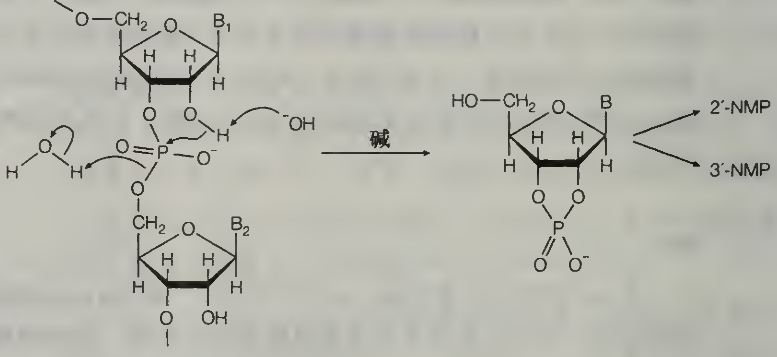


图 C2.2 碱性条件下 RNA 磷酸二酯键的分子内裂解

化学变性

一些化学物质能够使 DNA 或 RNA 在中性 pH 下变性，其中最熟知的例子是**尿素**（H_2NCONH_2）和**甲酰胺**（$HCONH_2$）。只要浓度稍高（几摩尔），这些物质便可破坏水溶液中的氢键作用。这就意味着，由堆积的疏水碱基形成的核酸二级结构在能量上的稳定性被削弱，则核酸变性。

黏性

细胞内的 DNA 分子很细长，术语称为高**轴比**（axial ratio）。DNA 的直径约 2nm，长度则达到微米（μm）、毫米（mm）量级，某些真核染色体甚至长达几厘米。打个比方，若 DNA 直径与意大利通心粉相当，则大肠杆菌（*E. coli*）的染色体长度（4.6×10^6 bp）约为 1km 长。另外 DNA 是一个比较刚性分子，其刚性程度几乎与半煮熟的意大利通心粉相似。DNA 的这些性质使得其水溶液具有**高黏性**（high viscosity）。另一方面很长的 DNA 分子又易被**机械力**（shearing forces）或**超声波**（sonication）损伤，同时黏度下降。当要完整分离大片段 DNA 分子，易被机械力破坏确是值得注意的问题；而超声波则可用来产生特定长度的 DNA 片段（参见 D4）。值得注意的是，无论机械力还是超声波均不能使 DNA 变性，仅仅

是降低水溶液中双链分子的长度。

浮力密度

可根据 DNA 的密度来对其进行纯化和分析。在高浓度高分子质量的盐溶液中，如 8mol/L **氯化铯**（CsCl）中，DNA 具有与溶液大致相同的密度，约为 $1.7 g \cdot cm^{-3}$。将溶液高速离心，则 CsCl 趋于沉降于底部，从而建立一密度梯度（图 C2.3）；而 DNA 最终沉降于与其浮力密度相应的位置，形成狭带。这种技术称为**平衡密度梯度离心**（equilibrium density gradient centrifugation）或**等密度梯度离心**（isopycnic centrifugation）（参见 A2）。在这种条件下，RNA 沉淀于试管底部而蛋白质浮于溶液表面，是一种 DNA 与 RNA 和蛋白质相分离的有效方法（参见 G2），而且由于 DNA 的精确的**浮力密度**（ρ）与 G+C 的含量呈线性关系，这也是一种有效的用于 DNA 分析法：$\rho = 1.66 + 0.098\%$（G+C）。因此，DNA 的沉降可用来确定其平均（G+C）含量；有时还可用来分离总体序列中具有不同（**G+C**）**含量**的片段（参见 D4）。

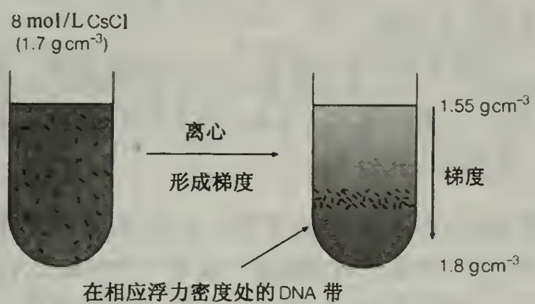


图 C2.3 DNA 的密度梯度离心

C3 核酸的光谱学和热力学特性

要 点

紫外光吸收	核酸中的芳香族碱基在 260nm 处具有最大光吸收。
减色性	核酸碱基的消光系数与它所处的环境有关。单一的核苷酸的吸收值比 RNA 和单链 DNA 的吸收值都大，单链 DNA 又比双链 DNA 大。这种双链 DNA 相对于单链 DNA 吸收值减少的现象就称为减色效应（hypochromicity）。
核酸定量	核酸在 260nm 处的光吸收可用于测定其浓度。当浓度为 1mg·ml^{-1}，光程为 1cm 时，双链 DNA 的 $A_{260}=20$，RNA 和单链 DNA 的 $A_{260}=25$。对于 RNA 和单链 DNA，A_{260}的值取决于碱基的组成和二级结构。
DNA 纯度	A_{260}/A_{280}比值可用于估计双链 DNA 样品的纯度。对于纯 DNA，该比值为 1.8；若比值大于 1.8，则意味着 RNA 污染；若比值小于 1.8，则有蛋白质污染。
热变性	升高温度可使 DNA 和 RNA 变性。RNA 随着温度升高而逐渐变性，双链 DNA 在某一确定温度“熔解”为单链 DNA，这一温度表示为 T_m，该值是 DNA 的（G+C）含量的函数。核酸的变性可以 A_{260}的变化来测定。
复 性	降温可使 DNA 复性，但仅当这一过程进行得足够缓慢时，才能使互补链退火而形成完全的非变性双链 DNA。
相关主题	核酸结构（C1）　基因组复杂度（D4） 核酸的理化特性（C2）

紫外光吸收　核酸因含有共轭的苯环而吸收紫外光，糖-磷酸骨架对光吸收的贡献并不明显。DNA 和 RNA 的最大紫外光吸收波长为 260nm（$\lambda_{max}=260$nm），这可以方便地与蛋白质的 λ_{max}（280nm）相区别（参见 B1）。核酸的光吸收特性可用来对其进行检测、定量和纯化。

减色性　尽管 DNA 或 RNA 的 λ_{max}为一常数，但其**消光系数**（extinction coeffecient）与碱基所处的环境有关。比较而言，单一的核苷酸在 260nm（A_{260}）的光吸收值最大，其次是单链 DNA（ssDNA）或 RNA，而双链 DNA（dsDNA）最小，这是由于碱基在疏水环境中的堆积所造成的（参

见 C1 和 C2)。这种光吸收的变化称为**减色性**，即 dsDNA 相对于 ssDNA 是**减色的**（hypochromic），反之 ssDNA 相对于 dsDNA 是**增色的**（hyperchromic）。

核酸定量

一般不太容易讨论核酸的摩尔消光系数（ε），因为其值与相应的分子长度有关。所以消光系数常以浓度 $mg \cdot ml^{-1}$表示：$1mg \cdot ml^{-1}$的 dsDNA 的 A_{260}为 20，相应地，ssDNA 和 RNA 的 A_{260}为 25。ssDNA 和 RNA 的 A_{260}值大致由以下两个因素决定：A_{260}的数值是所有不同碱基对光吸收的总和(嘌呤的消光系数比嘧啶大)，因此与分子中碱基的组成相关；dsDNA 因具有相同数量的嘌呤和嘧啶，这种效应并不明显。另一方面由于减色性，光吸收值也取决于给定分子的二级结构（双链区域）的数量。对于二级结构最少的寡核苷酸，通常由单链中每个单一的核苷酸的数值的总和来计算其消光系数。

DNA 纯度

dsDNA 的大致纯度（参见 G2）可由 260nm 和 280nm 处的光吸收的比值（A_{260}/A_{280}）来确定。与消光系数相同，光吸收谱（图 C3.1）的形状也会随着碱基的环境变化而变化。纯 dsDNA 的 A_{260}/A_{280}为 1.8，纯 RNA 的 A_{260}/A_{280}为 2.0，而蛋白质由于 λ_{max} = 280nm，A_{260}/A_{280}小于 1(实际上在 0.5 左右)。于是，如果 DNA 样品的 A_{260}/A_{280}大于 1.8，则说明有 RNA 污染；如果 A_{260}/A_{280}小于 1.8，则说明样品中含有蛋白质。

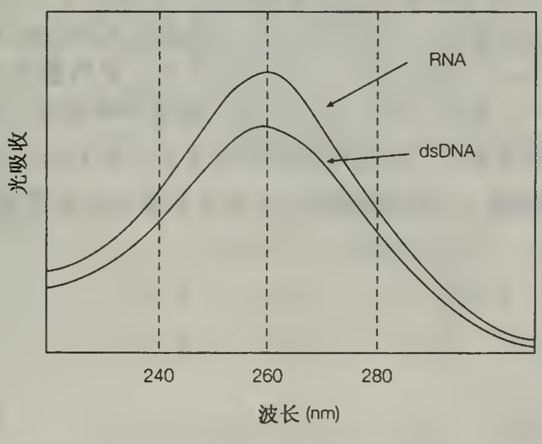


图 C3.1 同一浓度（$mg \cdot ml^{-1}$）的 RNA 和双链 DNA 溶液紫外吸收光谱

热变性

一些化学物质能够使核酸变性（参见 C2），加热同样能够导致 DNA 和 RNA 中双链部分氢键的破坏。双链核酸变性为单链核酸的过程可以方便地从光吸收的增加上加以观察（图 C3.2）。

dsDNA 与 RNA 的热力学表现不同，随着温度的升高 RNA 中双链部分的碱基堆积会逐渐地减少，其光吸收值也逐渐地、不规则地增大。较短

的碱基配对区域具有更高的热力学活性，因而与较长的区域相比变性快。而 dsDNA 的**热变性**或称**熔解**是**一协同过程**。分子末端、以及内部更为活跃的富含 A-T 的区域的变性将会使其附近的螺旋变得不稳定，从而导致整个分子结构在一确定温度的共同变性。这一温度取过渡期的中点值，称为**解链温度**（melting temperature，T_m）（图 C3.2a）。随着核酸的熔解，其光吸收值增大约 40%。T_m 是 DNA 样品中 G+C 含量的函数，对于长 DNA 分子，其取值范围是 80～100℃。

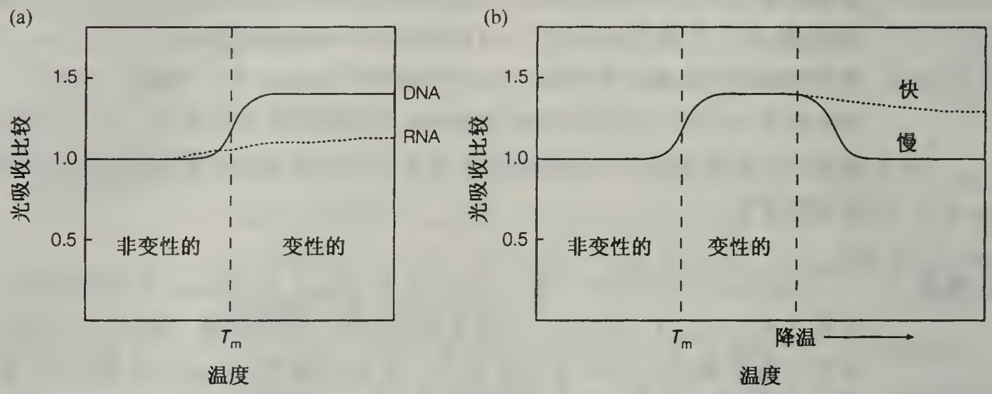


图 C3.2 (a) 双链 DNA 和 RNA 的热变性；(b) 快速和慢速降温下的 DNA 复性

复性

DNA 的热变性可以通过冷却溶液的方法复原。这一方法中，降温的速度对结果有一定影响。快速降温只形成局部区域的双链 DNA，经碱基配对或者在 DNA 内部或 DNA 之间短的互补区域的**退火**而形成，因此 A_{260}的下降很小（图 C3.2b)。另一方面，慢速冷却提供了足够的时间使得互补链能够相互配对，样品可以形成完全的双链 DNA，具有与原初样品同样的光吸收值。不同核酸链之间的互补部分的复性称为**杂交**（hybridization)。

C4 DNA 超螺旋

要 点

闭环 DNA	自然条件下，DNA 通常以闭环形式存在，即两条单链均为环状且相互连在一起。相互连接的数目称为连接数（linking number, *Lk*）。
超 螺 旋	超螺旋是 DNA 轴线的再次螺旋，若以 $Lk°$表示松弛闭环 DNA 的连接数，$Lk°$的改变将导致超螺旋。大部分天然 DNA 呈负的超螺旋，即 DNA 变形的方向与双螺旋解旋的方向相反。
拓扑异构体	具有特定连接数的环状 dsDNA 分子称为拓扑异构体（topoisomer)，除非其中的一条链或两条链均遭到破坏，其连接数不会发生改变。
缠绕和扭曲	超螺旋在几何上分为缠绕（twist）和扭曲（writhe)，前者指双螺旋是紧绕或松绕的，后者沿双螺旋轴本身的缠绕变化。根据等式：$\Delta Lk = \Delta Tw + \Delta Wr$，缠绕和扭曲是可以相互转化的。
嵌 入 剂	嵌入剂（intercalator)，如溴化乙锭（ethidium bromide，EB)，通过嵌入到配对碱基之间的方式与 DNA 结合，导致 DNA 双螺旋的局部解旋（untwisting)；如果是闭环 DNA，则会使缠绕增加。
超螺旋能	负超螺旋的 DNA 具有很强的扭力，这一能量促使 DNA 双螺旋的解旋，并驱动需解旋的 DNA 的解旋过程。
拓扑异构酶	通过暂时破坏 DNA 骨架的单链或双链，拓扑异构酶（topoisomerases）改变 DNA 超螺旋的水平。第Ⅰ类酶使 *Lk* 改变 ±1；第Ⅱ类酶使 *Lk* 改变 ±2。DNA 促旋酶（gyrase）利用 ATP 水解提供的能量为 DNA 引进负超螺旋，而拓扑异构酶Ⅳ使子染色体解开。
相关主题	原核生物的染色体结构（D1）　真核生物的染色体结构（D3） 染色质结构（D2）　细菌 DNA 复制（E2）

闭环 DNA　细胞内的许多 DNA 分子都是**闭环**双链，如细菌质粒、染色体和许多病毒 DNA 分子。这意味着不仅每条互补链自身的 3′端与 5′端连接成环状，而且两条互补链间相互螺旋缠绕。这样的分子中没有游离端，两条链缠绕的数目即为双螺旋的螺旋数，称为**连接数**（*Lk*）。

超螺旋 许多特性均起因于 DNA 分子的环状限制，理想方法是将 DNA 双螺旋想像成一个橡胶管，沿长径画一直线，以便追踪它的扭曲，将橡胶管的两端连接在一起形成环状（图 C4.1）。如果 DNA 双螺旋先扭曲，然后两末端相连，则这种变形被固定下来（图 C4.1）。这种被固定下来的变形称为**超螺旋**。超螺旋是 DNA 双链轴线的高度卷曲，与简单的环状相比，连接数有所变化。如果 DNA 扭曲方向与双螺旋方向相同，这种扭曲发生在闭合前，则形成的超螺旋定为**正**（positive）；反之定为**负**（negative）。平均来看，几乎所有细胞中的 DNA 分子超螺旋都是负的，即使对于像真核生物染色体这样依赖蛋白质支架形成大环的线性 DNA 分子也是如此（参见 D2 和 E3）。

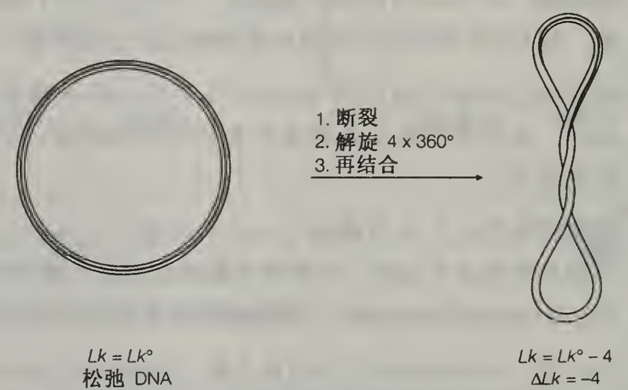


图 C4.1 DNA 超螺旋的橡胶管模型（图示了 DNA 构型上的变化和连接数的变化）

超螺旋的水平可以用连接数的变化（ΔLk）来衡量，ΔLk 即与**松散**时的闭环分子（Lk°）相比 Lk 发生的变化。这定义为闭环前 DNA 分子扭曲 360°的圈数。刚刚从细胞内分离出来的 DNA 分子通常是负的超螺旋，约 6 圈超螺旋/100 圈双螺旋（1000bp），表示为 $\Delta Lk/Lk^{\circ} = -0.06$。

拓扑异构体 闭环 DNA 的连接数是一个**拓扑学的**量，即除非 DNA 的一条或两条链发生断裂，否则这个量的数值不会发生改变。一个具有给定连接数的 DNA 分子就是一个**拓扑异构体**，各个拓扑异构体之间的区别仅在于它们的连接数不同。

缠绕和扭曲 当连接数不变时，DNA 分子的（几何）构型可以发生改变。一个超螺旋 DNA 拓扑异构体的两种极端构型可以用图 C4.2 来表示，分别代表了 ΔLk 全部参与形成**扭曲**（writhe）（图 C4.2a）或全部参与形成**缠绕**（twist）（图 C4.2c），橡胶管模型上的直线有助于观察 DNA 轴线的扭曲。介于这两种极端状态之间的平衡态（图 C4.2b），对应着扭曲和缠绕同时

发生改变的情况，ΔLk 在二者之间的分配可以用以下的方程表示：

$$\Delta Lk = \Delta Tw + \Delta Wr$$

其中 ΔLk 必须为一整数，而 ΔTw 和 ΔWr 则无此限制，它表明：一个超螺旋 DNA 分子的拓扑学改变被分配给了扭曲和（或）缠绕的改变而导致的分子构型的变化。

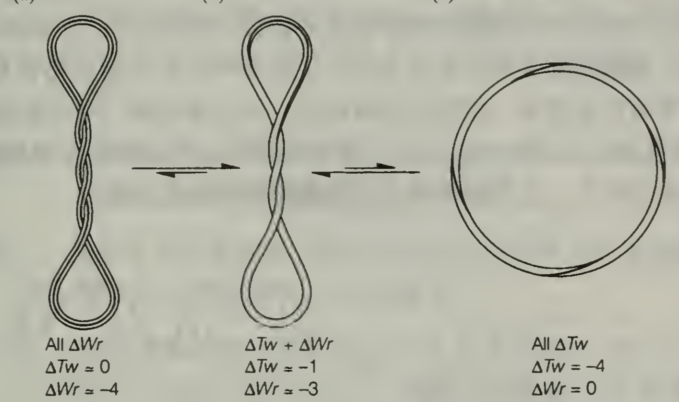


图 C4.2　用橡胶管模型表示的超螺旋 DNA 的构型可变性。在连接数不变的情况下，缠绕数和扭曲数的变化

嵌入剂

任何影响 DNA 双螺旋内部缠绕的因素都会使超螺旋分子的几何构型发生变化。例如：升温会减少缠绕，提高离子强度使缠绕增多，而另一个十分重要的因素就是**嵌入剂**的存在，其中最为人们所熟知的是**溴化乙锭**（图 C4.3a）。这是一种带正电的多环芳香族化合物，靠其自身嵌入到配对碱基之间而与 DNA 相连（图 C4.3b）。溴化乙锭的嵌入，不仅使其自身所发荧光大大增强，这也是用来对凝胶中的 DNA 进行染色的基础（参见 G4），而且使局部的双螺旋解旋约 26°。这意味着缠绕的减少，在闭环分子中会导致扭曲的增多。对于负超螺旋分子，就是（负）扭曲减少，其分子构型会沿图 C4.2 中 a→b→c 的方向而变化。

超螺旋能

超螺旋涉及向 DNA 分子引入了**扭转应力**，因此超螺旋 DNA 分子比松散的分子具有更高的能量。对于负超螺旋，这一能量会使 DNA 螺旋更易于局部解缠或解旋。因此负超螺旋有利于一切需要解旋过程的进行，诸如**转录起始**或**复制**（参见 E1 和 K1）。

拓扑异构酶

能够调控 DNA 分子超螺旋水平的酶称为**拓扑异构酶**。为了改变 DNA 的连接数，拓扑异构酶必须暂时断裂 DNA 分子的一条或两条链，通过攻击磷酸骨架上的酪氨酸残基，使自身与 DNA 的一端形成**磷酸酪氨酸键**以建立临时的共价连接。共有两类拓扑异构酶：**I 型**拓扑异构酶在 DNA 的一股链上产生一个切口，使另一条链得以穿越，连接数每次改变 ±1（图

C4.4a)；而Ⅱ型酶由 ATP 的水解提供能量，则在 DNA 的双链上产生切口，使另一双链 DNA 片段得以穿过，连接数每次改变 ±2（图 C4.4b）。大部分拓扑异构酶会降低正或负超螺旋的水平，也就是沿着能量最小的方向作用。然而，**DNA 促旋酶**，一种细菌中的Ⅱ型酶，利用 ATP 水解提供的能量向 DNA 分子引入负超螺旋，从而抵消了 DNA 复制中产生的正超螺旋（参见 E2）。由于这一作用机制中包含一个双链穿过另一双链的过程，Ⅱ型拓扑异构酶还能够被用来拆开 DNA 分子，如复制中产生出的连在一起的新生分子（参见 E2）。在细菌中，该功能由**拓扑异构酶Ⅳ**担当。对所有的生物体，拓扑异构酶都是一种必要的酶，涉及复制、重组和转录（参见 E1、F4 和 K1）。DNA 促旋酶和拓扑异构酶Ⅳ是**抗细菌药物**的作用目标，而Ⅰ，Ⅱ型酶都是人类**抗癌制剂**的作用目标。

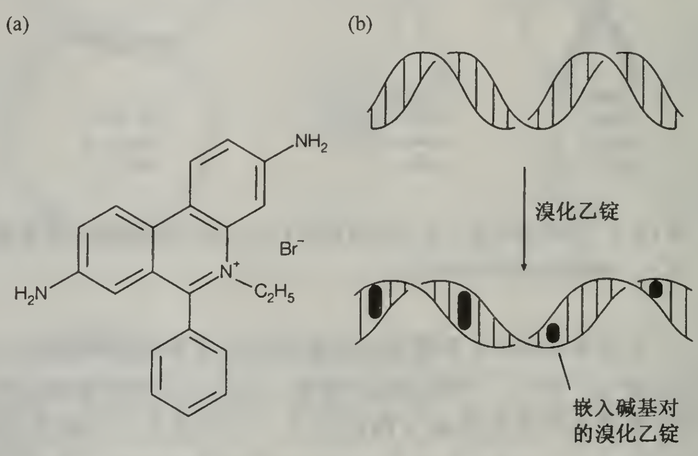


图 C4.3 (a) 溴化乙锭；(b) 嵌入过程，图示了 DNA 双螺旋的增长和解缠

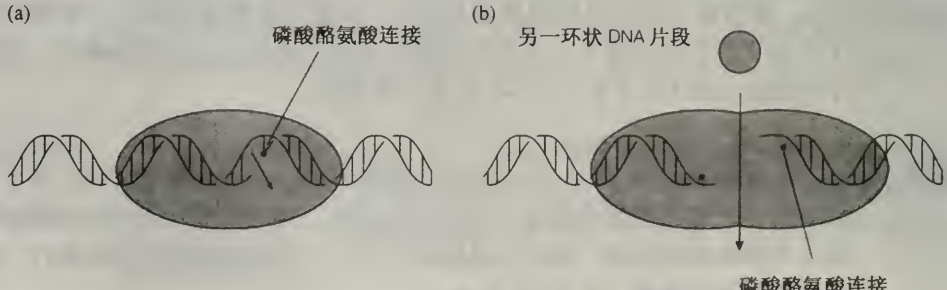


图 C4.4 拓扑异构酶Ⅰ型 (a) 和拓扑异构酶Ⅱ型 (b) 的作用机制（详细见正文）

（王薛林 译 刘进元 校）

D1 原核生物的染色体结构

要 点

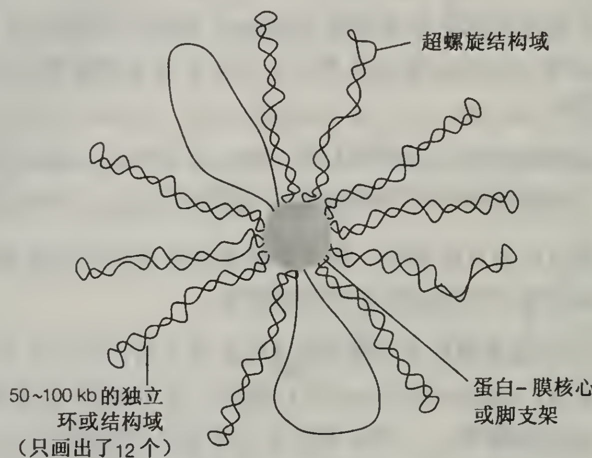
大肠杆菌染色体	大肠杆菌染色体为闭环 DNA，长度约 4.6Mb，集中分布在称为拟核（nucleoid）的区域内。在正常生长情况时，DNA 保持持续复制。
DNA 结构域	基因组由 50～100 个大环，或长度为 50～100kb 的结构域组成，与膜蛋白复合体（membrane-protein complex）结合而被固定。
基因组超螺旋	基因组为负超螺旋，每个结构域在拓扑学上是独立的，结果是各结构域可保持不同水平的超螺旋。
DNA 结合蛋白	DNA 结构域被非特异性的 DNA 结合蛋白如 HU 和 H-NS（类组蛋白，histone-like protein）缠绕，这些蛋白质束缚着约半数的 DNA 超螺旋。一些其他的分子如宿主整合因子（integration host factor）、RNA 聚合酶及 mRNA 等均有助于类核的组构。
相关主题	DNA 超螺旋（C4） 基因组复杂度（D4） 染色质结构（D2）

大肠杆菌染色体 原核生物的基因组可用大肠杆菌（*E. coli*）染色体为例来加以说明，其由长为 4.6Mb 的单一闭环 DNA 分子组成（参见 C4），被包裹在称为**拟核**的区域内。该区域内 DNA 浓度高达 30～50mg·ml^{-1}，并包含着所有与 DNA 结合的蛋白质如聚合酶（polymerase）、阻抑物（repressor）和其他成分（参见下文）。在试管中 DNA 浓度 1mg·ml^{-1}就是相当高了。在正常生长时，DNA 保持持续复制；当生长速率达到最大时，平均每个细胞中含有基因组的两个拷贝（参见 E2）。

DNA 结构域 小心地将大肠杆菌的 DNA 与它的大多数结合蛋白分离开，在电子显微镜下就可观察到拟核的组构，其 DNA 由 50～100 个**环**或**结构域**组成，这些环或结构域的末端被与细胞膜一部分相连的蛋白而固定（图 D1.1）。环的大小为 50～100kb，目前还不能确定这些环究竟是静态（static）还是动态（dynamic），但有一种模型暗示在这些环的基部 DNA 缠绕于聚合酶或其他酶作用位点处。

基因组超螺旋 尽管有证据表明单个结构域呈单独的超螺旋，但就整体而言，大肠杆菌的染色体是负超螺旋（$\Delta Lk/Lk^{\circ} = -0.06$，参见 C4）。电子显微镜表

明有一些结构域不形成超螺旋，这可能是因为DNA的一条链已断裂（参见C4），而另一些结构域清晰地含有超螺旋（图D1.1）。DNA与膜蛋白支架的结合可能阻碍了DNA的自由旋转，因而结构域可能在拓扑学上是独立的。然而还没有真正的生化证据证实细胞内不同区域的结构域的超螺旋水平存在明显的差异。



图D1.1 电子显微镜下*E. coli*染色体（4600kb）结构概图，细线表示DNA双螺旋结构

DNA结合蛋白

染色体中环形的DNA结构域由于大量的DNA结合蛋白相互作用而进一步受到束缚，这些蛋白质中最多的是HU蛋白以及一种分子质量较小的碱性（带正电）二聚体蛋白，非特异地与其周围的DNA结合，**H-NS**蛋白（以前称为**H1蛋白**）是一种中性的单体蛋白，也与DNA不同序列非共价结合，但是倾向与DNA内部弯曲区域结合。这些蛋白质有时被称为**类组蛋白**（histone-like protein）（参见D1），具有压缩DNA的作用，这对DNA被包装进入类核以及稳定和限制染色体超螺旋是十分必要的。虽然被分离的染色体$\Delta Lk/Lk^{\circ}=-0.06$，大约每17圈DNA螺旋存在一个超螺旋，其中有半数永久性地被周围的蛋白质所**束缚**如HU（实际上呈扭曲型，参见C4），而另一半则没有被束缚，可以适应不同的扭曲和缠绕（参见C4）。研究发现RNA聚合酶、mRNA分子以及位点专一性DNA结合蛋白［如宿主整合因子（IHF），一种HU的同系物，与特定的DNA结合导致DNA弯曲140°］对于DNA结构域的构成都是很重要的。尽管高度有序的DNA-蛋白复合物诸如核小体（nucleosome）（参见D2）还未被检测到，类核的组构是相当复杂的。

D2 染色质结构

要 点

染色质	真核生物染色体都包含一个很长的线性 DNA 分子，并被组装在细胞核(nucleus)内。染色质用来描述构成真核生物染色体的高度有序的 DNA-蛋白质复合物。染色质结构被用来包装和组构染色体 DNA，并能在细胞周期的不同阶段调整压缩 DNA 的水平。
组蛋白	染色质中的主要蛋白质成分是组蛋白：一类分子质量较小的、碱性（带正电）二聚体蛋白，与 DNA 紧密结合。组蛋白有四个家族：H2A、H2B、H3 和 H4，以及另一类具有不同特性、特定功能的 H1。不同物种具有不同的组蛋白变异体。
核小体	核小体核心是染色体结构的基本单位，由组蛋白八聚体构成即含四种核心组蛋白各两分子。146bp 的 DNA 以左手方式环绕八聚体 1.8 圈。环绕在核小体上的 DNA 占据了真核生物 DNA 的几乎所有的负超螺旋。
H1 的功能	一个单一 H1 分子在出入核小体核心颗粒处对 DNA 起稳定作用，并在核小体间分配 DNA。包括 H1 的核小体核心颗粒加上 H1 称为染色小体（chromatosome）。某些情况下，H1 可被结合得更紧密的 H5 所代替，并与无转录活性的 DNA 结合。
连接 DNA	核小体核心颗粒间的连接 DNA 长度从几个 bp 到大于 100bp 不等，但通常约 55bp。这样核小体重复的单位长约 200bp。
30nm 纤丝	染色质被组构成 30nm 纤丝或螺线管的巨大结构体，每圈大约由 6 个核小体左手螺旋组成。大多数染色质以这种形式存在。
高级结构	在最大规模上，染色体 DNA 以 30nm 纤维的形式构成长达 100kb 的环，这样的环被蛋白支架和核基质所固定。整体结构与原核生物 DNA 的结构域有些类似。
相关主题	DNA 超螺旋（C4）　真核生物的染色体结构（D3） 原核生物的染色体结构（D1）　基因组复杂度（D4）

染色质　　真核生物细胞 DNA 的总长度因不同的种而异，但与原核生物基因组相比，要大数千倍，并由一定数目的分散的**染色体**组成（人类为 46 条）。每条染色体中的 DNA 为单一线性分子，长度可达几厘米（参见 D3）。所有这些 DNA 分子必须被包装进大小相当于细菌细胞的细胞核内（参见

A2)；实际上，DNA 以高度浓缩的形式存在，其浓度可高达 $200mg \cdot ml^{-1}$（参见 D1)。高度有序的 DNA-蛋白质复合体，是一种核蛋白复合体即**染色质**的形成完成了这一紧密组装的目的（参见 A4)。蛋白质组分占染色质质量的 50%以上。在细胞周期的不同进程，染色体极大地改变着它们的包装水平（参见 D3 和 E3)：从**中期**（细胞分裂期）的高度浓缩的结构到**间期**的极为松散结构，这些都暗示出染色质的不同组构水平。

组蛋白

真核生物染色质中的蛋白质主要是**组蛋白**，共分五类或五个家族，即核心组蛋白：H2A、H2B、H3 和 H4，以及 H1。**核心组蛋白**的分子质量较小，约 10～20kDa；H1 稍大，约 23kDa。所有组蛋白带有大量正电荷，序列中 20%～30%由碱性氨基酸——赖氨酸及精氨酸组成（参见 B1)。这意味着在形成染色质时组蛋白将与带负电荷的 DNA 紧密结合。

同一类组蛋白在不同的物种之间相当保守，如植物和动物之间，这显示出组蛋白对染色质的重要作用。在特定物种中，某类组蛋白通常存在几种非常相似的变异体，会分别在不同的发育阶段、不同的组织中表达。在不同类的组蛋白之间没有序列的相似性，但结构研究显示，它们却具有相似的三级结构（参见 B2)，暗示出所有组蛋白在进化上是相关联的（参见 B3)。

组蛋白 H1 在某些方面同其他几类组蛋白是不同的：除了分子质量较大以外，在不同种之间或同种不同小种之间在序列上比其他几类组蛋白具有更多的差异。H1 更容易从染色质中提纯，可能仅为其他几类组蛋白存在量的一半，而其他几类组蛋白在染色质中的量相差不多。这些事实表示出 H1 在染色质结构中具有特殊的、独特的作用。

核小体

19 世纪 70 年代的一些研究主要集中在染色质结构的基本单位。核酸酶是一种能够水解核酸的磷酸二酯键的酶。外切酶从核酸链末端释放单个的核苷酸，而内切酶则作用于链内部的磷酸二酯键。用**微球菌核酸酶**（一种切割双链 DNA 的内切酶）处理染色质，会分离长度为 200bp 的不同倍数的片段。研究发现，每个 200bp 的片段都与由 4 种组蛋白各 2 分子即 $(H2A)_2(H2B)_2(H3)_2(H4)_2$ 构成的**八聚体**核心紧密结合，以及 1 分子 H1 松散结合，这就是之所以被称为核心组蛋白的原因。这些蛋白质保护 DNA 免受微球菌核酸酶的作用。继续反应将导致 H1 的丢失，并产生抗性极强的与组蛋白八聚体结合的 **146bp** 片段。这样的结构称为**核小体核心**(nucleosome core)。不论染色质来源于何处，都具有相似的核小体结构。

通过 X 射线晶体学的研究，对核小体核心颗粒的结构在细节上有了更为深入的了解。组蛋白八聚体形成楔形盘，146bp 长的 DNA 以左手方式环绕其周围 1.8 圈。图 D2.1 表示了这种结构的基本特性和空间构型。这种围绕核小体的 DNA 左手环绕形成负超螺旋，即**超螺旋转折**（superhelical turn)（术语称为扭曲，参见 C4)。与原核生物相比，虽然真核生物

DNA 具有相似水平的负超螺旋，但实际上几乎所有超螺旋都是环绕核小体的结果，并没有不被约束的超螺旋（参见 D1）。

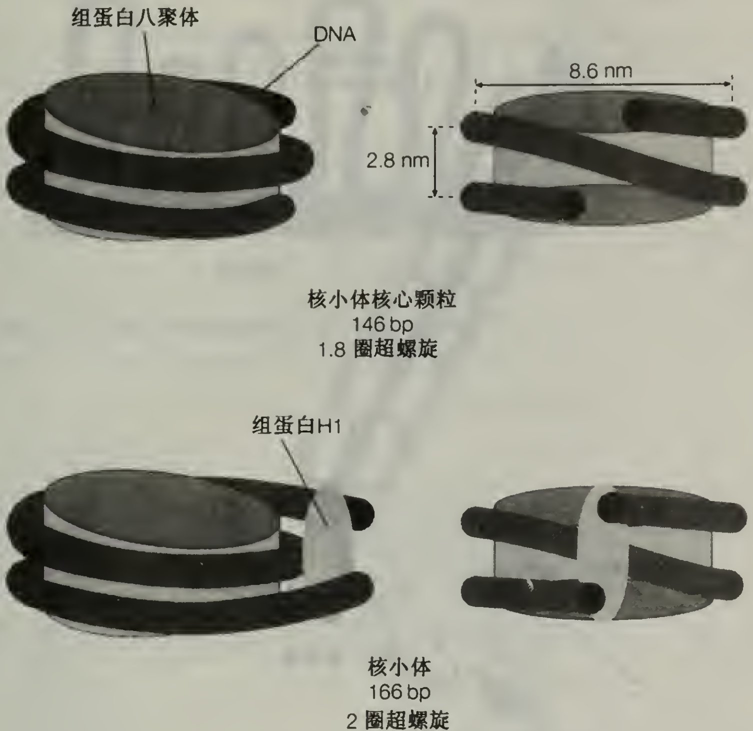


图 D2.1 核小体核心颗粒和染色小体的结构概图

H1 的功能

1 分子的组蛋白 H1 与核小体结合，在 DNA 出入核小体核心颗粒处对 DNA 起稳定作用（图 D2.1）。由于 H1 的存在，又有 20bp 的 DNA 得到保护而免受核酸酶的降解，总体上受保护的 DNA 达到 166bp，相当于绕组蛋白八聚体两圈。核小体核心颗粒加上一个 H1 分子构成**染色小体**(chromatosome)。与核心组蛋白相比，H1 分子要大一些，是因为有一个以稳定核小体核心之间 DNA 的 C 端末尾。综上所述，H1 在序列上比其他组蛋白具更大的变异；在有些细胞类型中，H1 被一种极端变异的组蛋白 H5 所取代，H5 与染色质紧密结合，与不进行转录的 DNA 相连。

连接 DNA

在电子显微镜下，可以观察一定条件下的 DNA 核小体核心颗粒呈一种称为“**串珠**（beads on a string)”的结构。这一结构包含球形颗粒（核小体），颗粒间由很细的 DNA 链相连。**连接 DNA** 是在微球菌核酸酶实验中重复出现的，形成 200bp 片段所需的额外 DNA 部分（参见上文）。核心颗粒间的连接 DNA 的平均长度为 55bp，但在不同的种或不同的组织中其长度会在几乎等于 0 到多于 100bp 的范围内变化（图 D2.2）。

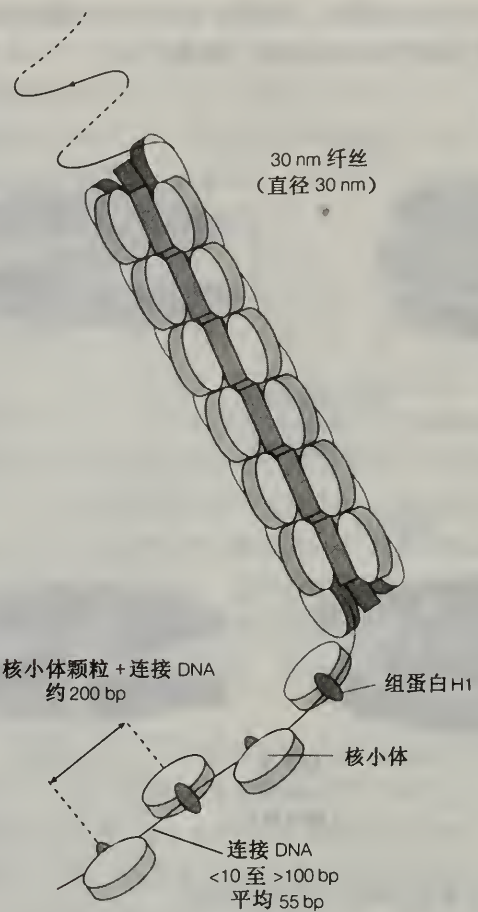


图 D2.2　由连接 DNA 相隔核小体链和 30nm 纤维示意图

30nm 纤丝

组蛋白 H1 的存在加强了“串珠”的组构，在电子显微镜下呈现一种锯齿形的结构。随着盐浓度的改变，核小体进一步形成直径为 30nm 的**纤丝**。详细的研究表明这是核小体被组装成高度有序的左手螺旋，又叫**螺线管**（solenoid）的过程，每圈螺旋约 6 个核小体（图 D2.2）。然而仍有一些关于纤丝结构精细组构的猜测，包括连接 DNA 所采取的路径以及不同长度的连接 DNA 是以何种方式被组装成看来十分一致结构的，大部分体内染色体 DNA 被装配成 30nm 纤丝（参见 D3）。

高级结构

染色质的最高水平组构与原核生物 DNA 十分相似（参见 D1）。电镜下观察除去组蛋白的染色体，可以发现与 D1 中图 2.1 相似的环状结构域。真核生物染色体具有更多的环，但这些环的大小与原核生物基本相同，约为 100kb DNA。这些环被称作**核基质**（nuclear matrix）蛋白质复合体所束缚（参见 E3），环中的 DNA 以 30nm 纤丝的形式存在，构成了幅

宽约 300nm 的排列（图 D2.3）。

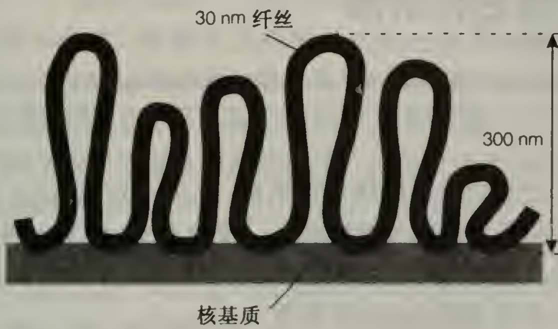


图 D2.3 30nm 纤丝形成染色体环结构示意图

D3 真核生物的染色体结构

要 点

有丝分裂染色体	有丝分裂时配对姐妹染色单体（sister chromatid）的经典图片展示出染色质的高度浓缩状态。线性 DNA 沿着一条单一路径从染色体一端到另一端，以 30nm 纤丝构成的长达 100kb 的连续环的方式锚定在核心处的核基质上。
着丝粒	着丝粒是两条染色单体相连的区域，也是通过动粒与纺锤体连接的部位。在分裂后期纺锤体将姐妹染色单体向细胞两极拉开。着丝粒具有特别短的 DNA 序列特征，在哺乳动物中属于卫星 DNA。
端粒	线性 DNA 的末端形成端粒结构来保护 DNA 分子免受降解和渐进的截短。端粒是由一种专一端粒酶合成的短重复序列构成，独立于正常 DNA 复制。
间期染色体	在间期染色体的环状结构虽然与核基质相连，但采用了一种比较松散的结构。
异染色质	异染色质是中期染色质的一部分，结构比较紧密，无转录活性。
常染色质	常染色质是中期染色体比较松散的区域，包括无活性的 30nm 纤丝区和转录活跃区，在转录活跃区中纤丝已被解散，单个核小体也可能被转录起始蛋白所替代。
DNase Ⅰ 超敏性	染色质活性区，或因结合特异蛋白或因进行转录而被松散的 30nm 纤丝的区域，其特征是对 DNA 酶Ⅰ（DNaseⅠ）的高度敏感性。
CpG 甲基化	哺乳动物 DNA 中 5′-CG-3′序列通常在胞嘧啶碱基处被甲基化；但在频繁转录基因的启动子附近产生不被甲基化的 CpG 岛（island），形成对 DNaseⅠ高度敏感的区域。
组蛋白变异体和修饰	染色质的浓缩程度受下列条件控制，至少部分受其控制，一是对组蛋白的化学修饰，以改变其细胞周期中它们所带的电荷数；二是在特定的细胞类型中，或在不同的发育阶段通过利用组蛋白变异体来实现。

相关主题

染色质结构（D2）　　真核生物的 DNA 复制（E4）
基因组复杂度（D4）

有丝分裂染色体

下面的几幅示意图实际上表示染色体**有丝分裂**中高度浓缩的状态（图 D3.1）。细胞分裂时，子染色体被**纺锤体**拉向细胞两极，如果不是呈这种紧密的结构，脆弱的几厘米长的染色体将会被所产生的拉力扯断。图 D3.1 所示的结构真实地展示出由同一染色体复制产生的两个**姐妹染色单体**在**着丝粒**处相连，染色体的末端是**端粒**，即 DNA 分子的末端。虽然 DNA 采取相当复杂的螺旋路径，但总的来说是以线性的方式排列在染色体的长轴上。

图 D3.1 也显示了有丝分裂染色单体的横截面结构。染色体的环状结构（参见 D2）将由蛋白质组成的中心支架或核基质展成扇形，这些环是由 30nm 纤丝组成的染色质，一种可能性就是这些连续的环在染色体的长轴方向上以螺旋状的路径排列。

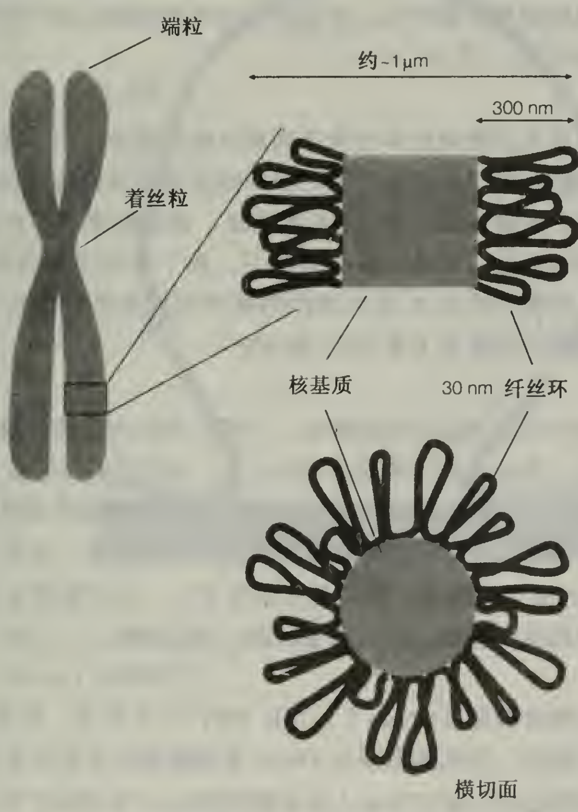


图 D3.1 有丝分裂染色体的结构图

着丝粒

着丝粒是在分裂中期两条姐妹染色单体相连的紧缩区域，也是**动粒**——一种与纺锤体中的**微管**（microtubule，参见 A4）相连的蛋白复合体的组装场所。微管在**分裂末期**行使将染色单体分开的功能。酵母 DNA 的着丝粒仅由 88bp 长的富含 AT 的序列组成，与两段很短的保守区域相连；而在哺乳动物中，着丝粒是由相当长的序列构成，其侧翼是大量的重

复序列，即**卫星 DNA**（参见 D4）。

端粒

端粒是形成真核生物染色体线性 DNA 分子末端的特化了的序列，由数以百计的短重复序列（人类中是 5′- TTAGGG - 3′）构成，这些序列是**端粒酶**（如**核糖核蛋白**，ribonucleoprotein，参见 O1）以独立于正常 DNA 复制的机制合成的。端粒 DNA 可以形成一种特殊的二级结构，其功能是保护染色体末端免遭降解。端粒的自主合成起到阻止正常的复制不能拷贝线性 DNA 分子的末端造成染色体渐渐变短的作用（参见 E4）。

间期染色体

在分裂间期，染色体上的基因被转录，DNA 被复制（在 S 期，参见 E3 和 E4）。此时，在大半个细胞周期染色体均呈一种非常松散的结构，因而不能观察到单个染色体，但染色体的环状结构应该仍然存在，与核基质相连（参见 D2，图 D3.5）。

异染色质

异染色质是间期染色体内保持高度浓缩的部分，但其紧密程度比中期时要差一些。在电子显微镜下能观察到细胞核四周较浓的区域即异染色质区，可能是由紧密压缩的 30nm 纤维构成。最近越来越多的证据表明异染色质没有转录活性。尽管在某些情况下，整个染色体都呈异染色质状态，如雌性哺乳动物的两条 X 染色体，仍可以相信异染色质的大部分是由靠近染色体着丝粒的重复卫星 DNA 构成的。

常染色质

染色质中的非异染色质的部分，习惯上统称为**常染色质**，是全部转录发生的区域。常染色质的构成是不均一的，其中有由 30nm 纤丝构成的染色体环是相当不活跃的区域，以及基因正在被活跃转录或将要被转录的区域（约占总体的 10%），这里 30nm 纤丝已被解离成“串珠”结构（参见 D2）。这些区域中的部分，特别是在启动子内，已去除核小体结构以结合转录因子和其他一些蛋白质（图 D3.2）（参见 M5）。

DNase Ⅰ 超敏性

脱氧核糖核酸酶（DNaseⅠ）可以切割 DNA 骨架，除非 DNA 因结合蛋白质而被保护，因此染色质对 DNaseⅠ的敏感性可被用来定位细胞中具转录活性的染色质。对 DNaseⅠ超敏感的较短区域被认为那里的 30nm 纤丝已被序列特异性调节蛋白的结合所破坏（图 D3.2），结果裸露出 DNA 链，从而很易受 DNaseⅠ的攻击。较长的敏感区域代表着那里的转录正在进行。这些区域随细胞类型不同而异，同时与那些细胞中特异表达基因的位点相关。

CpG 甲基化

在哺乳动物细胞中，一种可能传递信号使得表达基因位点处的染色体保持适当的包装水平的重要化学修饰是序列 5′- CG - 3′中对胞嘧啶 C - 5 的甲基化即通常所称的 **CpG 甲基化**。在哺乳动物细胞中 CpG 位点通常被

甲基化，因而在整个基因组中含量相对较少，这是由于 5 -甲基胞嘧啶（5 -methylcytosine）可以自发地脱氨（deaminate）生成胸腺嘧啶，而这种错误又往往得不到修复，所以甲基化的 CpG 很快地突变为 TpG（参见 F1 和 F2）。CpG 的甲基化与染色质中的非转录区相连，而在整个基因组中存在着未甲基化的 CpG 的“**岛**”，其 CG 二核苷酸的比例比一般水平高得多。这些岛通常的长度约为 2000bp，并且恰好与 DNase Ⅰ 敏感区域相吻合（图 D3.3）。CpG 岛包围着在几乎所有细胞类型中都表达的**持家基因**（housekeeping gene）的启动子区域，可能大部分游离出核小体。

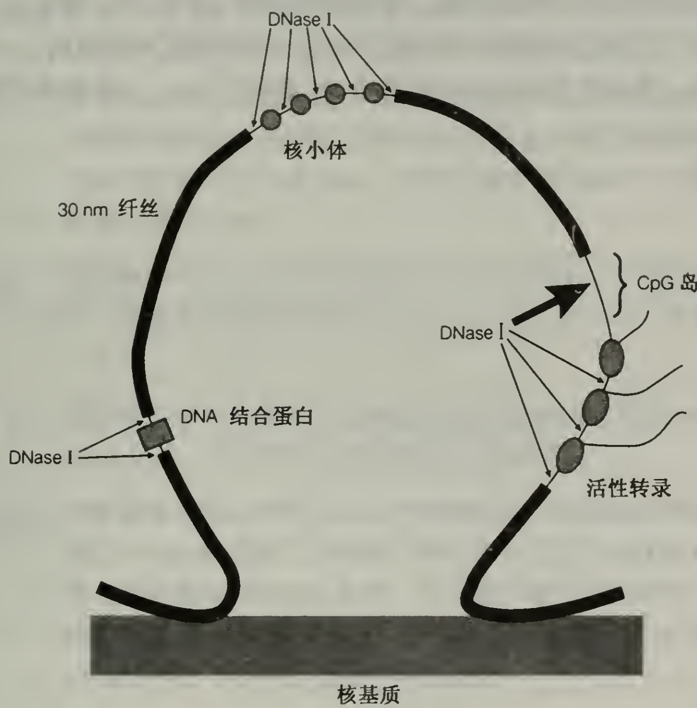


图 D3.2 常染色质，用粗细箭头表示出活性与非活性区域（详见正文）以及 DNase Ⅰ 超敏感位点

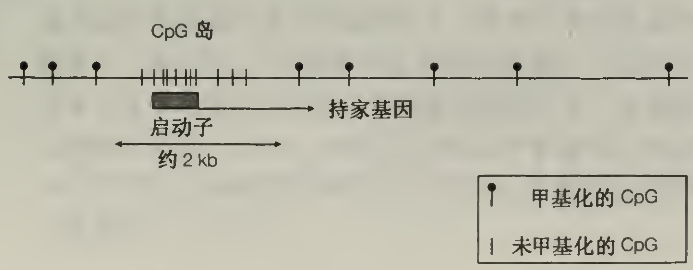


图 D3.3 持家基因的 CpG 岛及其启动子

组蛋白变异体和修饰

有关染色质的压缩和解压缩的主要机制被认为是通过执行包装的组蛋白来操纵的（参见 D2)。细胞周期中被包装的染色质的快速变化可通过对组蛋白的化学修饰来进行调控，例如活跃转录的染色质与核心组蛋白（参见 D2）的 N 端赖氨酸（lysine）残基的乙酰化（acetylation）（参见 B2）相关联，而有丝分裂中染色体的浓缩伴随着组蛋白 H1 的磷酸化。这些变化改变了组蛋白所带的正电荷，会影响染色质构象的稳定性，其构象主要指 30nm 纤丝或纤丝间的相互作用。

染色质浓缩的慢速变化则与不同的发育阶段和不同的组织类型有关。这些变化与可变的组蛋白变异体的利用（参见 D1）相联系，也许同样是改变了染色质构象的稳定性。组蛋白 H5 是这种效应的一个很好的范例，在一些非常不活跃的染色质中 H5 取代了 H1，如在禽类红细胞中（参见 D2)。

D4 基因组复杂度

要 点

非编码 DNA

真核生物细胞的大部分 DNA 并不编码蛋白质；大部分基因含有较长的内含子（intron）序列，基因或基因簇（gene cluster）被大段未知功能的序列所分割。这样的 DNA 由许多有些相关的短序列元件的重复组成，有时重复次数高达数百至数千万个。

复性动力学

这是根据序列的拷贝数不同在基因组 DNA 复性过程中的复性速率不同的原理，来识别不同种类的重复序列 DNA 的一种技术。用这一技术可将人类 DNA 分为高度重复 DNA、中度重复 DNA 和单一序列 DNA。

单一序列 DNA

复性最慢的 DNA 组分就是单一序列 DNA，一般由单一拷贝基因，或重复仅数次的基因组成。*E. coli* 的 DNA 基本上都是单一序列。

串联基因簇

基因组中串联重复数百次的某些基因或基因簇区域构成中度重复 DNA 区段，如 rRNA 基因和组蛋白基因簇。

分散重复 DNA

中度重复 DNA 中的大部由数百碱基对（SINES，即短散布元件）或 1～5000 碱基对（LINES，即长散布元件）的 DNA 序列构成，每个序列重复 100 000 多次，并分散于整个基因组中。在人类基因组中最突出的例子就是 *Alu* 和 *L1* 元件，这些可能是寄生 DNA 序列，自身可以通过转座复制。

卫星 DNA

卫星 DNA 多位于染色体的着丝粒附近，可能与有丝分裂纺锤体的附着相关，由大量串联重复短序列（多至 30bp）组成。卫星 DNA 的高变性是 DNA 指纹技术的基础。

遗传多态性

基因或染色体基因座上的碱基变化（突变）能使该基因座产生多种形式（多形体），这就被称为基因的多态性。这个术语可以描述单个个体中单拷贝基因的不同突变遗传因子。普通类型包括单核苷酸多态性（SNP）和简单序列长度多态性（SSLP），其中 SNP 可创造或破坏限制酶的识别序列，结果产生了限制性片段长度多态性（RFLP）。

相关主题

核酸的光谱学和热力学特性（C3）
RNA 聚合酶 I 基因：核糖体重复（M2）

非编码 DNA

复杂的真核生物基因组 DNA 是原核生物如 *E. coli* 的 1000 多倍。显然其中大部分 DNA 不编码蛋白质，因此人体中的蛋白质种类并不是 *E. coli* 的 1000 倍。基因的编码区被**内含子**中断（参见 O3），基因可以长达数千碱基对，但在基因组中基因并不紧密排列，而是被很长的未知功能的序列隔开。这些**非编码** DNA 的大部分是由不同类型序列的相似或相同多重拷贝组成。这些拷贝可以是一个接一个依次排列（**串联重复的**）如**卫星 DNA**，或者可以是多拷贝散于整个基因组中（**分散的**）如 ***Alu* 元件**。

复性动力学

在对基因组序列结构大规模测序之前（参见 J2），基因组序列复杂性的研究是通过测量变性 DNA 的复性速率来进行的（参见 C3），用超声波破碎或机械力将基因组 DNA 剪切成大小大致相同（从几百至几千碱基对）的片段（参见 C2），加热变性后在低浓度下复性，在一定时间内（t）DNA **复性**的比例是由初始 DNA 浓度（C_0）决定的。但在基因组中有多个拷贝的单链片段比单一序列有更多的选择性，可以与许多互补链结合，结果复性速度比单一序列快得多。DNA 链的复性速率可以用光谱法测量（参见 C3），或用更灵敏的分离单链与双链的**羟磷灰石层析法**测量。用剩余的单链 DNA 的比例（f）对 C_0t 作图，得到的曲线代表了 DNA 样品的**复性动力学**，这就是通常所称的 **C_0t 曲线**。

图 D4.1 显示了理想状态下人类和 *E. coli* 基因组 DNA 的 C_0t 曲线模式，可以看出，人类 DNA 复性分为三个不同阶段（1～3），相对于 DNA 拷贝数的降低，或**复杂度**的增加。阶段 1 对应于**高度重复 DNA**（＞10^6 拷贝/基因组），阶段 2 对应于**中度重复 DNA**（＜10^6 拷贝/基因组），阶段 3 对应于**单一序列 DNA**（一个或几个拷贝/基因组）。实际上这样划分有些专断，相互间并没有明显的分界，*E. coli* 的 DNA 复性仅有一个阶段。

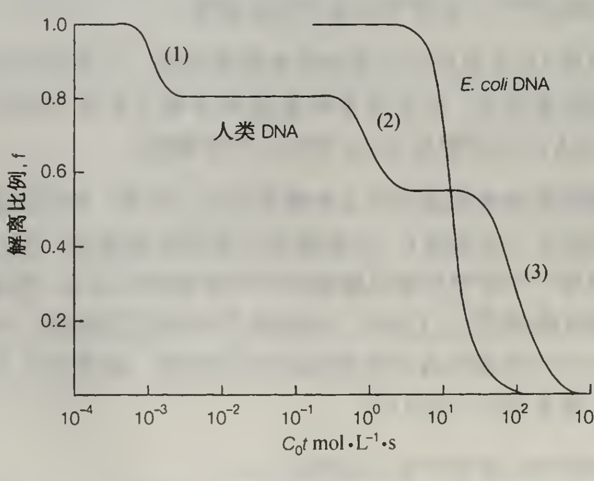


图 D4.1　人类和 *E. coli* 的 DNA 的理想复性动力学曲线模式图

单一序列DNA

基因组中的这部分 DNA 在 $C_0 t$ 曲线中处在最慢的复性阶段，对应于基因的编码区，在单倍体基因组中通常只有一个或几个拷贝以及那些单一的间隔序列。在 *E. coli* 基因组中，实际上所有 DNA 都是单一序列，主要是由大体上相连的单拷贝基因组成。但是 *E. coli* 的 DNA 只是人类细胞 DNA 的千分之一，在任一给定的浓度下，任一特定序列都有相对较多的可选择序列，其复性较快，其 $C_0 t$ 值也相对较小。

串联基因簇

中度重复 DNA 由许多类型的重复序列组成。多重复序列组成的基因簇在重复区域的下游出现，这些通常是其产物需求量很高的基因，**rDNA**（编码 rRNA）就是其中一例。编码 18S、5.8S 和 28S rRNA 的 45S 前体的基因就有约 10 到 10 000 个拷贝的重复**排列**，其拷贝数因物种而异（参见 M2，图 D4.1）。人类基因组中 45S 基因在 5 个不同的染色体上都有排列，每一排列包含约 40 个拷贝。在分裂间期，这些区域都处于**核仁区**，核仁是细胞核中的致密区域，是 rRNA 合成和修饰的工厂（参见 M2 和 O1）。另一个串联基因簇的例子是组蛋白基因，其产物在 S 期大量产生。5 个组蛋白基因同在一个**基因簇**，在某些生物中，可直接重复达数百次。

分散重复DNA

许多物种的中度重复 DNA 大部分是由数百（SINES，即**短散布元件**）至 1000～5000 个（LINES，即**长散布元件**）碱基序列重复上千次而构成，并分散于整个基因组中，称为**分散重复 DNA**。*Alu* 元件是人类基因组中最常见的分散重复序列，即 300bp 的 DNA 序列重复 300 000 至 500 000 次。这些拷贝的 80%～90%是相同的，绝大部分包含 *Alu* Ⅰ 限制位点（参见 G3），故称为 *Alu* 元件。*L1* 元件是另一个分散序列，*L1* 和 *Alu* 元件几乎占了人类基因组的 10%，存在于基因间以及内含子中。

这些分散重复 DNA 具有很多潜在的功能，从复制起始点（参见 E4）到基因调控序列（参见 N）。也许最重要的功能就是 *Alu* 元件和其他这样的基因家族很可能可以转座方式在整个基因组中随机地复制自身（参见 F4）。这些序列可以是寄生的，或者是自在的，也许其 DNA 序列自身没有功能，但通过影响或中断基因序列，在进化中起着很重要的作用。

卫星DNA

真核生物基因组中高度重复 DNA 由一些 2bp 至 20～30bp 的极短序列，以数千个拷贝的串联方式排列。这种排列又称为**卫星 DNA**，是因为在染色体 DNA 片段的 CsCl 密度梯度离心中，由于浮力密度的不同，在主带 DNA 附近出现的卫星带。因为由很短的重复序列组成，其 G+C 的含量也不一般（参见 C2）。根据重复序列的长度不同，卫星 DNA 可分为**小卫星 DNA**（**不确定数目串联重复，VNTR**）和**微卫星 DNA**（**简单串联重复，STR**），微卫星 DNA 由最短的重复序列组成。图 D4.2 显示了一个果蝇的卫星 DNA 序列，该序列在果蝇基因组中出现几百万次。与分散重复序列一样，卫星 DNA 还没有被证明有任何功能。其中那些集中在染色体

着丝粒附近，并形成一大片异染色质区的卫星 DNA，很可能起着着丝粒与动粒结合的作用。小卫星 DNA（VNTR）一般出现在染色体的末端，而微卫星 DNA 则较为平均地沿染色体分布。

用来鉴定个人与其家族的亲缘关系的 DNA 指纹技术（DNA fingerprinting technique）的基础就是小卫星 DNA 的重复序列。在某些卫星 DNA 序列的排列中拷贝数是**高度可变的**（hypervariable），这在不同个体中的变异极大。这些变异体现了基因的多态性（见下面及 F1）。通过限制性酶切、southern 印迹和杂交（参见 J1）可以确定基因组中几个不同卫星 DNA 排列组中的 DNA 的精确长度，进而可以鉴别一个特定个体。也就是说，另一个个体（除了双胞胎或同一个克隆）具有同样一套排列的长度的概率是极其微小的。因为这种排列是从双亲那里各得一半，亲缘关系也可以通过比较不同个体间的长度来确定。

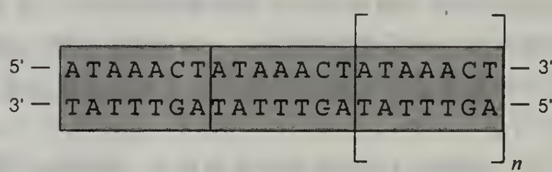


图 D4.2 果蝇的卫星 DNA 重复单位

遗传多态性

如果同一种类的不同染色体或不同个体之间的某一条染色体上的相同区域或基因座上含有两个（或更多）稍微不同的 DNA 序列，这就被描述为多形体，而该基因座则被称为展示了（遗传）多态性。基因多态性是由突变导致的（参见 F1）。譬如在一个双倍体个体中，一个特定的单拷贝基因上的两个突变遗传因子可以因为仅仅一个核苷酸不同而不同，从而该基因就被称为具有多态性。相似地在一个群体中，如果同一个基因存在其他可选择的 DNA 序列，肯定会有许多编码该基因的不同的等位基因（在整个群体中）。导致基因多态性的突变事件（参见 F1）不仅会导致如上所述的**单核苷酸多态性**（SNP），而且会导致数组重复序列在长度上的变化。重复的每一不同长度都是一种不同的形式，也是**简单序列长度多态性**（SSLP）的例子。

SNP 可以发生在限制酶识别（参见 G3）的短序列中，因此用限制酶消化一个 DNA 分子所产生的片段长度对每一个等位基因都是不相同的。这就是所谓的**限制性片段长度多态性**（RFLP）。如果一个特定的 RFLP 与一个遗传疾病的等位基因相关联，它就可以在临床诊断上用来做标记。RFLP 还可以用来帮助画出染色体的遗传图，因此可以帮助搞清基因组测序计划（参见 J2）中的 DNA 序列的顺序。可以用凝胶电泳（参见 G3）检测出像限制性片段（参见 G3）或 PCR 产物（参见 J3）这样的具有相同长度的片段中的 SNP。在此过程中，DNA 片段必须先变性，这样才能使

两个分开的链在电泳迁移中保持只依赖于其核苷酸序列的构象。这种技术被称为**单链构象多态性**（SSCP）。

D5 遗传信息流

要 点

中心法则	中心法则最初的观念就是 DNA 到 RNA 到蛋白质，其中转录与翻译是分别进行的。虽然有很多例子与其有相矛盾的地方，但它还是在大范围内是正确的。反转录病毒能将 RNA 反转录为 DNA，很多病毒能直接以 RNA 为模板复制 RNA，还有很多生物能编辑信使 RNA 序列，因此蛋白质编码序列不仅仅直接由特异的 DNA 序列决定。
原核基因表达	单个基因或操纵子的转录从启动子开始，结束于终止子，产生单顺反子或多顺反子的信使 RNA。核糖体从起始密码子（靠近核糖体结合位点）到终止密码子翻译信使 RNA 的编码区。转运 RNA 根据遗传密码把适当的氨基酸递送到延长的蛋白链上。
真核基因表达	大多数情况下转录起始于启动子，单顺反子信使 RNA 从一个基因转录而来。产生出的前信使 RNA 在 5′端有帽子结构，在 3′端有多聚 A 尾巴。在成熟 mRNA 从核中释放到细胞质中被核糖体翻译之前，其中的内含子通过剪接被去除。
相关主题	原核生物的转录（K）　遗传密码与 tRNA（P） 真核生物的转录（M）　蛋白质合成（Q） RNA 加工与核糖核蛋白复合体（O）

中心法则

在 20 世纪 50 年代早期，Francis Crick 提出遗传信息沿着从 DNA 到 RNA 到蛋白质单向传递，即 DNA 产生 RNA，RNA 产生蛋白质。这就是著名的分子生物学的**中心法则**，虽然在当时它的每一步还没有被足够证据所证明。我们现在知道中心法则的基本框架是正确的，虽然我们还有必要对基本框架作出很多修正。有关遗传信息传递的更完整的图形表述如图 D5.1 所示，基本的路线仍然是从 DNA 到 RNA 到蛋白质。在原核和真核生物细胞中，DNA 被**转录**（参见 K）为 RNA 分子（**信使 RNA**），它包含了相同的序列信息（虽然 U 代替了 T），然后这些信息按**遗传密码**（参见 P1）**翻译**为（参见 Q）蛋白序列。我们把 DNA 复制（参见 E）也包括在图 D5.1 中，通过对 DNA 分子信息的复制形成两个子代 DNA 分子。

但是，也确认出这个基本框架的一些例外。许多病毒包含由 RNA 分子构成的基因组（参见 R4）。**反转录病毒**，包括引起 AIDS 的致病病毒 **HIV**，单链 RNA 分子被转换为双链 DNA 拷贝（T 代替 U），随后就插入到宿主细胞基因组。该过程被称为反转录，其相应的由病毒基因组编码的

酶，称之为**反转录酶**。还有很多病毒的 RNA 基因组直接被复制为 RNA，而不利用 DNA 作为中介（**RNA 复制**），例如冠形病毒和丙型肝炎病毒。到目前为止还没有发现蛋白质能特异性地决定特定的 RNA 或 DNA 序列的例子，因此中心法则的翻译步骤确实表现出单向性。

在一些生物中，发现了 RNA 在从 DNA 转录后被改变的例子，这个过程被称为 **RNA 编辑**（参见 O4），所以 DNA 序列不是非常忠实地被用于编码产物蛋白。

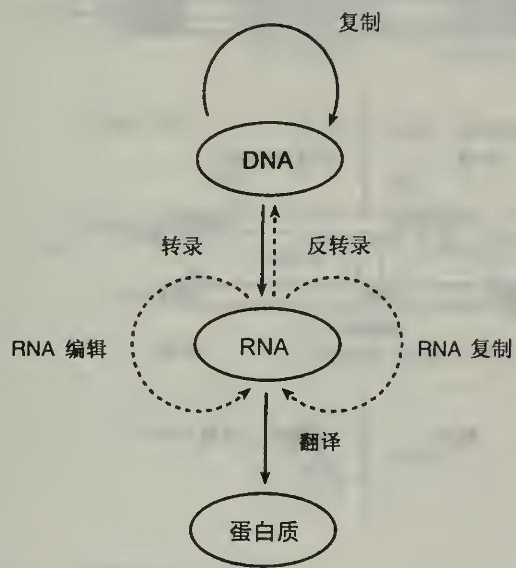


图 D5.1 遗传信息的传递

原核基因表达

图 D5.2 和图 D5.3 表示出与真核与原核细胞**基因表达**的相似点和不同点。在原核生物中（图 D5.2），**RNA 聚合酶**（参见 K2）转录基因起始于**启动子**，这是 DNA 上一段特异性的启动转录（参见 K3）的序列，终止于它特异性地终止转录的**终止子**（参见 K4）。产生的 mRNA 可能包含编码一个或多个蛋白的区域。在后一种情况下，信使 RNA 被描述为是**多顺反子**，编码这部分的 DNA 区域称为**操纵子**（参见 L1）。mRNA 的编码区在**核糖体**（参见 O1 与 Q）上被翻译成蛋白质，该过程首先是核糖体在**核糖体结合位点**（RBS）与 mRNA 结合，通过**转移 RNA**（tRNA，参见 P2）按遗传密码运送氨基酸（氨酰- tRNA）从起始密码子（通常为 AUG）到终止密码子把信息翻译为氨基酸序列。

真核基因表达

在真核生物中，转录发生在核内而翻译则发生在细胞质中，因此 mRNA 具有更长的寿命。转录的**前 mRNA**（pre-mRNA）由三种 RNA 聚合酶中的一种（**RNA 聚合酶Ⅱ**，参见 M4）来合成，且通常只编码一个蛋白（即**单顺反子**）。RNA 在核中通过在 5′端加帽和 3′端加 **poly（A）尾巴**

增加其稳定性（参见 O3）。在大多数真核基因中，蛋白编码区被非编码的间隔序列（称为**内含子**）所打断。因此最终前 mRNA 中的内含子必须被去掉以产生连续的蛋白编码序列（**外显子**），这一过程称为**剪接**，是由一种被称为 **snRNP**（**核内小核糖核蛋白**，参见 O3）的 RNA-蛋白质复合所介导。成熟 mRNA 随即就被运出细胞核并在细胞中经核糖体翻译成蛋白质。

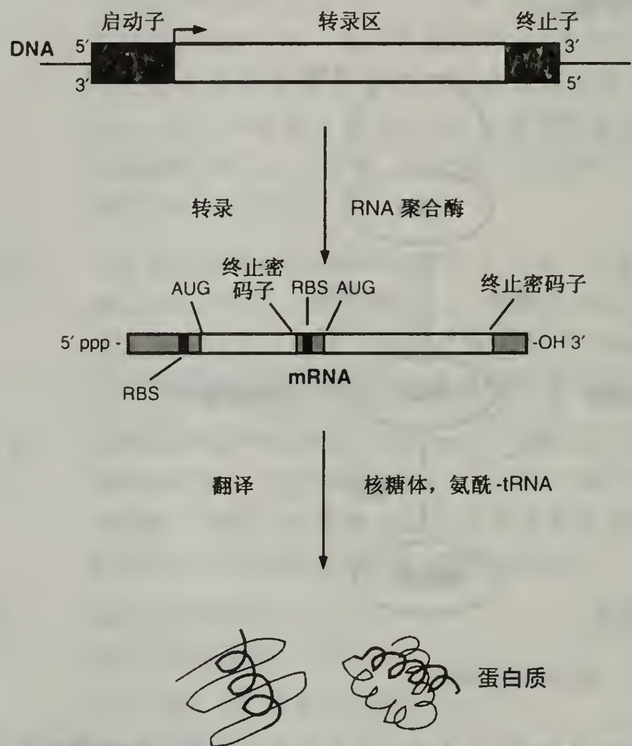


图 D5.2　原核基因表达简图

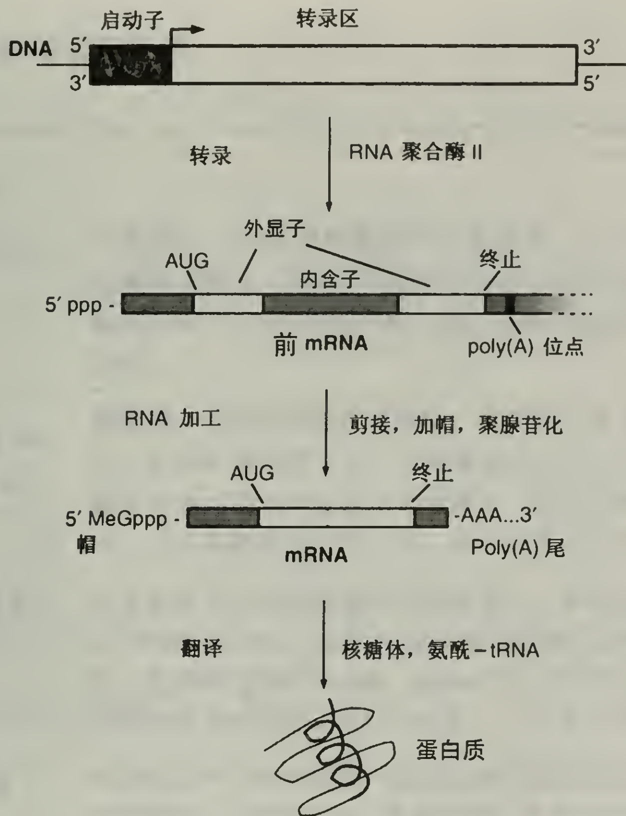


图 D5.3　真核基因表达简图

（王薛林 译　刘进元 校）

E1　DNA 复制概述

要　点

半保留机制	在复制中，DNA 双螺旋的两条链解开，并分别作为模板指导以脱氧核苷酸 5′-三磷酸为前体的互补子链的合成。这样每个子细胞接受亲代 DNA 双链中的一条。这种机制可用密度标记实验来证明。
复制子、复制起点与终点	像细菌、病毒这样的小染色体，是以单一复制子的单位进行复制的。复制从特定的位点——复制起点开始，双向进行直至终点。真核生物的染色体包含多个复制子，每个复制子有自己的复制起点，并在复制中融合成一体。起点处富含 AT，以便于解链。
半不连续复制	每个复制叉中的前导链是连续复制的，而后随链是以反方向合成不连续的短片段。这样的短片段在原核生物中约 1000～2000nt 长，在真核生物中约 100～200nt 长，由 DNA 连接酶连接。这种机制的存在是因为 DNA 只能由 5′ →3′方向合成。
RNA 引导	前导链及所有的后随链片段的 DNA 合成是都由短的 RNA 片段引导起始，然后沿 DNA 模板延伸。引物在连接前被去掉并填补上 DNA。这种机制有助于保持高度的复制保真性。
相关主题	原核与真核生物的染色体结构（D）　DNA 修复（F3） 细菌的 DNA 复制（E2）　噬菌体与真核生物病毒（R） 真核生物的 DNA 复制（E4）

半保留机制　DNA 复制机制的核心在于 DNA 双螺旋中的两条链都携带相同的信息，它们的碱基序列是互补的（参见 C1）。这样在复制中两条**亲代链**分开，分别作为**模板**指导酶催化的新生互补**子链**的合成，并遵循标准的碱基配对原则（A 与 T，G 与 C）。两条新生双链在细胞分裂时分别进入两个子细胞（图 E1.1a）。亲代链分开及新生 DNA 开始复制处称为**复制叉**。1958 年，Meselson 和 Stahl 用实验证实了**半保留**机制（图 E1.1b）。大肠杆菌细胞在含有^{15}N 标记的培养基中生长几代，它们的 DNA 就完全被^{15}N **密度标记**上（两条链均被标记：$^{15}N/^{15}N$）。然后将细胞转移到含有正常^{14}N 的培养基中，每一次细胞分裂后，都取样提取 DNA 并进行 CsCl **平衡密度梯度离心**，可根据浮力密度将不同的分子分开（参见 A2 与 C2）。第一代细胞分裂后，DNA 复制了一次，全部是杂合分子，在梯度上处于纯$^{15}N/^{15}N$ DNA 与纯$^{14}N/^{14}N$ DNA 之间。在^{14}N 的培养基上第 2 代细胞分裂

后，一半是杂合的$^{15}N/^{14}N$的密度，一半是纯合的轻链（$^{14}N/^{14}N$）。在接下来每一代中，$^{14}N/^{14}N$ DNA的比例逐渐上升，但仍有一部分是杂合的$^{15}N/^{14}N$分子。上述结果只能用半保留模型来解释。所有DNA分子都以这种方式复制，模板以3′→5′方向被读（参见C1），而新生链以5′→3′方向合成。

DNA合成的底物是脱氧核苷三磷酸（dNTP）：**dATP**、**dGTP**、**dCTP**和**dTTP**。其反应机制涉及由新生链3′端核苷酸的3′-OH与下一个碱基互补的dNTP的α-磷原子处的亲核攻击，结果是焦磷酸的除去。合成的能量来自于dNTP的水解，并产生焦磷酸。

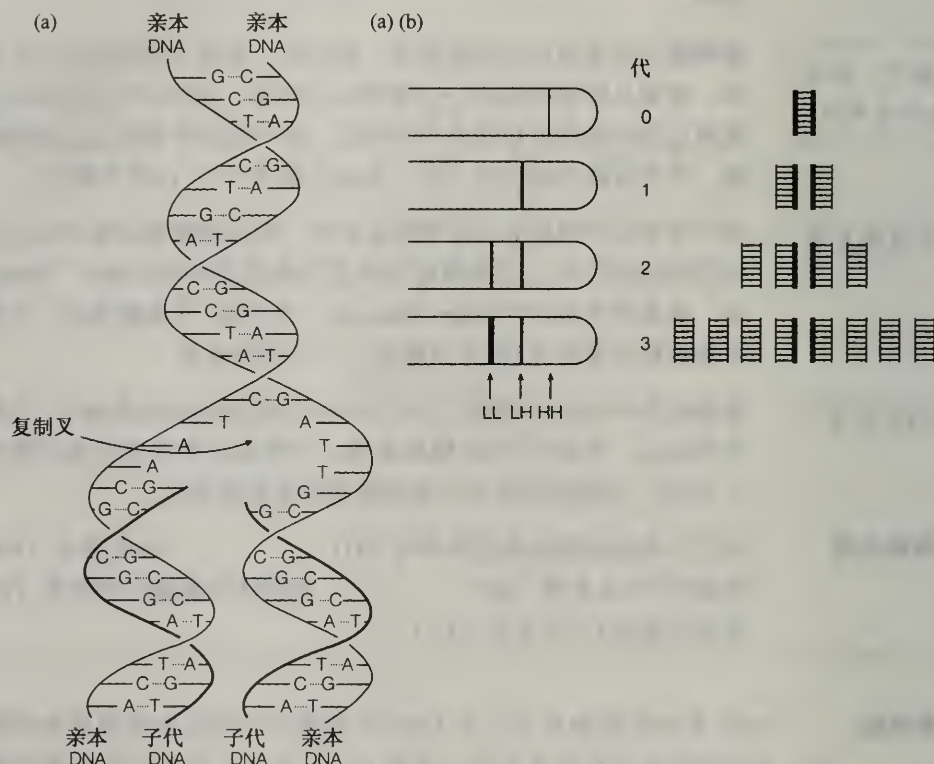


图 E1.1 （a）复制叉处的DNA半保留复制；（b）半保留复制的验证。左边的离心管表示在生长相应代后完全被^{15}N标记的DNA（HH）、杂合（LH）及完全被^{14}N标记（LL）DNA带的在密度梯度中的位置。右边的图表示相应的DNA分子。最初的^{15}N标记的DNA分子由粗线表示，所有^{14}N子链DNA均用细线表示

复制子、复制起始与终点

以单一单位复制的任一段DNA都称为**复制子**。在单个复制子内的DNA复制是从一被称为**起始点**的固定点起始的（图E1.2a）。通常两个复制叉从起始点向**两个方向**进行复制，随着双链的解链模板被复制直至**终点**。所有的原核生物染色体、许多噬菌体以及病毒的DNA分子都呈环形（参见D1），并由单一复制子构成，因此与惟一起点成约180°处就有1个

单一终点（图 E1.2a）。线性病毒 DNA 分子通常只有 1 个单一起始点，但不一定在该分子的中间。不管怎样，起点都是一个复杂的区域，DNA 复制的起始与生物生长周期调节都在这里进行并协调。相比之下，真核生物的线形染色体是由多复制子构成，每个复制子都有自己的起始点。一个典型的哺乳动物细胞有 50 000～100 000 个复制子，每个复制子长约 40～200kb。在相邻复制叉的**复制泡**相遇处，新生 DNA 融合并形成复制完整的 DNA（图 E1.2b）。真核染色体起始点序列要比单复制子 DNA 的复制起始点简单得多，因为真核 DNA 复制的调控主要是在 S 期的开始而不在每一单个起始点（参见 C3）。所有起始点都在双链最初解链处均富含 AT 序列，富含 AT 的起始点比富含 GC 的起始点更易解链（参见 C3）。

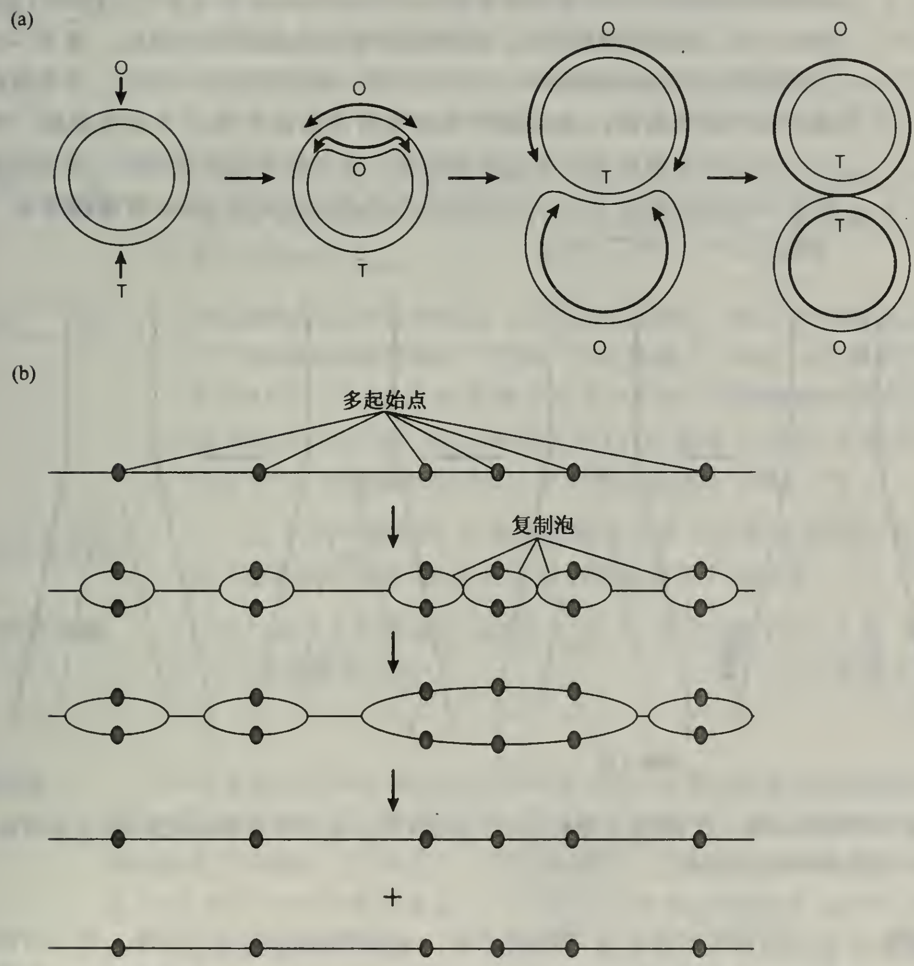


图 E1.2　(a) 一个环状细菌的复制子的双向复制。两个复制叉均从起始点（O）向背向终点行进（T），子链 DNA 用粗线表示；(b) 真核生物的多复制子，每一起始点均用黑点表示

半不连续复制

在半保留复制中，两条 DNA 的新生链合成在同一复制叉中同时进行。但 DNA 复制原则上只允许以 5′→3′方向进行。因两条 DNA 链是**反向平行**的，那么方向为 5′→3′的那一条亲本链又是如何被复制的呢？似乎应该是以 3′→5′方向进行，而事实上并非这样。如果把 *E. coli* 细胞用 ^{3}H 胸腺嘧啶（一个具放射活性的 DNA 前体）标记几秒钟，当用碱性蔗糖梯度对新生 DNA 进行离心，根据分子大小变性 DNA 的单链会分开，结果发现大量新合成的 1000～2000nt 长（真核中是 100～200nt 长）的小片段。如果这些细胞进而在未标记的胸腺嘧啶中培养，那这些**冈崎片段**很快便连接成大分子的 DNA。这些结果可以用**半不连续复制**模型来解释（图 E1.3）。一条链即**前导链**是从起始点按 5′→3′方向进行连续复制，而另一条链即**后随链**不立即复制而是被约 1000～2000nt 长（真核生物中约 100～200nt 长）的短片段取代的。后随链的复制从复制叉处起始，按 5′→3′的反方向朝向起始点形成第 1 个冈崎片段。随着复制叉的前进，前导链继续被复制成连续长链，而后随链片段按反方向以不连续方式被复制。事实上，后随链与前导链两者合成的物理方向与前导链是相同的，因为后随链在复制叉处回转了 180°。合成后不久冈崎片段就被 **DNA 连接酶**连成一条连续的 DNA（参见 E2）。

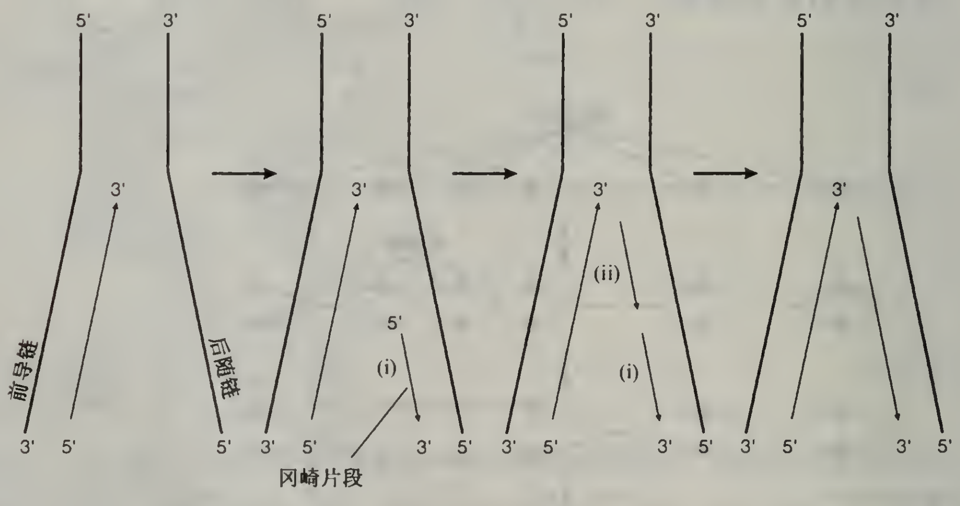


图 E1.3 半不连续复制。图中标出了第 1 个（i）及第 2 个（ii）冈崎片段或新生片段在后随链上的形成，只显示出单个复制叉

RNA 引导

对冈崎片段的仔细研究表明，与前导链的头几个一样，每一片段 5′端的头几个核苷酸均是核糖核苷酸，因此 DNA 合成是由 RNA **引导**的（参见 E2）。这些引物在片段连接之前被去掉，产生间隙由 DNA 填补。用 RNA 引导 DNA 复制的原因很可能是出于保证 DNA 复制的高**忠实性**（参见 F4）。

E2 细菌的 DNA 复制

要 点

实验系统 遗传上简单的噬菌体和质粒是用于体外研究较大而脆性的细菌染色体很有用的模型系统。这些最简单的 DNA 几乎完全依赖于宿主细胞的复制蛋白。大一些的噬菌体可编码一些它们自己的复制因子。

起 始 复制是通过起始的速率来调控的。在 *E. coli* 的复制起始点 *oriC* 处的起始涉及到在起始蛋白复合体（DnaA-ATP）处的 DNA 的缠绕以及在富含 AT 序列处双链的解链，然后解旋酶 DnaB 结合上去为复制而延伸单链区域。DnaA 酶的含量与生长速度相关。

解 旋 亲本链 DNA 被 DNA 解旋酶及单链结合蛋白作用而解旋，产生的正超螺旋被拓扑异构酶即 DNA 旋转酶所释放，旋转酶是抗生素作用的靶目标。

延 伸 移动着的引发体在后随链上合成多个引物，DNA 聚合酶Ⅲ全酶的二聚体延伸前导链及后随链，α 亚基聚合 DNA，ε 亚基校正新生 DNA 分子。DNA 聚合酶Ⅰ的 5′→3′外切核酸酶活性切去后随链上的 RNA 引物，聚合酶同时用 DNA 填补上 DNA 于缺口中，然后 DNA 连接酶将后随链上的冈崎片段连成一体。

终止与分离 *E. coli* 的两个复制叉在与复制起始点成 180°的复制终点相遇，相互连锁的子链在 DNA 拓扑异构酶的作用下相分离。

相关主题 原核与真核生物的染色体结构（D） 真核生物的 DNA 复制（E4） DNA 复制概述（E1） DNA 修复（F3）

实验系统 DNA 复制的大部分知识是利用制备 DNA 以及完成复制所必需的相关蛋白质及其因子构成的体外系统而获得的。但像 *E. coli* 这样的原核生物的染色体大而脆弱，不能用上述方法来研究。因而小且简单的噬菌体和质粒 DNA 被广泛用作模型系统。其中遗传上简单的噬菌体 *ΦX*174 为 *E. coli* 染色体复制的研究提供了最佳模型，因为它几乎完全依赖于细胞的复制因子来进行自身复制。*ΦX*174 的复制结构体是一个仅有 5kb 的超螺旋环状 DNA。大一点的噬菌体如 T7（40kb）和 T4（166kb）编码多个自身复制所需蛋白，并采用一些独特的机制来完成自身复制。虽然不是很好的模型，但仍能提供丰富的信息，且可为阐明复杂的 DNA 复制机制提供很多有用的解答。

起始

虽然细菌 DNA 复制也从复制起始点起始，但噬菌体及质粒并不是研究合适细菌 DNA 复制的模型，这是因为噬菌体及质粒 DNA 在单个细胞分裂周期可以复制许多次，而细菌染色体的复制与细胞生长周期紧密偶联。*E. coli* 的起始点位于遗传基因座 *oriC*，并与细胞膜相连。这一区域已被克隆并构建成类似 *E. coli* 的染色体的，且易于研究的**微型染色体**。*oriC* 包含 4 个 9bp 的起始因子 DnaA 蛋白的结合位点。起始蛋白的合成与生长速度相偶联，因而复制的起始也与生长速度相偶联。细胞高速生长期，原核染色体可以在第一轮复制结束之前就从两个新形成的起始点开始第 2 轮复制。在这种情况下，子细胞得到的是部分还在进行复制的染色体（图 E2.1）。一旦 DnaB 蛋白达到一个足够量，即形成由 30～40 个分子构成的复合体，每一个分子结合 1 分子 ATP，在其周围的 *oriC* DNA 呈缠绕形（图 E2.2）。该过程需要 DNA 形成负超螺旋（参见 C4）。负超螺旋有助于 3 个 13bp 长的富含 AT 的重复序列的解链，同时允许 DnaB 的蛋白结合。DnaB 是个 **DNA 解旋酶**。解旋酶是利用 ATP 水解的能量结合并解开双链 DNA（或 RNA）。用这种方法形成的单链泡被**单链结合蛋白**（Ssb）所覆盖，以防止发生断裂及重新复性。**DNA 引发酶**结合到 DNA 上并合成一段短的 RNA 引物，从而引发第一个复制叉的**先导链**的复制，接下来是双向复制。

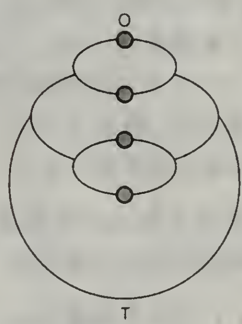


图 E2.1 第一轮复制完成前，细菌在新的起始点处（O）又起始了新一轮复制。T 表示终止点

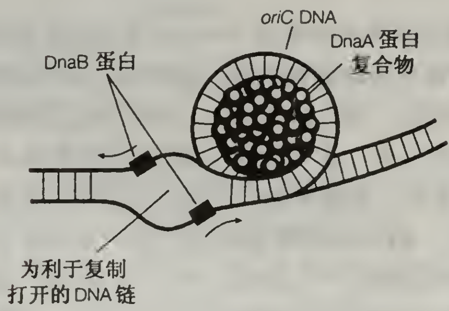


图 E2.2 通过 DnaA 蛋白周围的 *oriC* DNA 的缠绕打开起始点处 DNA 双链以利于复制

解旋

为了使起始于起始点的复制不断进行，DNA 解旋酶须沿着模板链前进，并打开双螺旋使复制顺利进行。除 DnaB 外，另一 DNA 解旋酶可结合到另一条链上以帮助解链。Ssb 蛋白的结合进一步推进了解链的进行。但在闭合环形 DNA 分子中，在复制叉处的螺旋转角的去除会导致分子其余部位额外转角的导入，结果形成了正超螺旋（参见 C4）。虽然环形 DNA 本身具有的负超螺旋部分地抵销了正超螺旋，但对于复制叉的不断前进仍是不够的。这些正超螺旋可通过**Ⅱ型拓扑异构酶**即 **DNA 旋转酶**作用引入负超螺旋而得到不断释放（参见 C4）。DNA 旋转酶的抑制剂如新生霉素（novobiocin）及奥啉酸（oxolinic acid）是细菌复制的有效抑制剂，也具有抗生素活性。

延伸

随着新形成的复制叉替代亲本链的后随链（参见 E1），1 个包含 DnaB 解旋酶和 DNA 引物酶，被称作**引发体**的移动复合体在后随链上每隔 1000～2000nt 就合成 1 个 RNA 引物。前导链和后随链的引物均由 **DNA 聚合酶Ⅲ全酶**来延伸。该多亚基复合体是一个二聚体，一个单体用于先导链的合成，另一个用于后随链的合成。1 个复合体中具有两个聚合酶可以保证两条链以同样速度进行合成。二聚体的两个单体都含有聚合酶 α 亚基，和一个具有 **3′→5′校正作用的外切核酸酶** ε 亚基。阅读校正有助于维持复制的高度忠实性（参见 F1）。β 亚基将聚合酶结合于 DNA。而在每一个单体中的其余亚基各不相同，以保证全酶分别在先导链上和后随链上合成长的或短的 DNA 片段。一旦后随链的引物被 DNA 聚合酶Ⅲ延伸，会由 **DNA 聚合酶Ⅰ**来将引物去除，并填补上 DNA。DNA 聚合酶Ⅰ具有 5′→3′聚合酶、5′→3′外切核酸酶以及 3′→5′校正外切核酸酶活性。5′→3′外切核酸酶在去掉引物的同时，聚合酶通过延伸邻近冈崎片段 3′端 DNA 来填补缺口（图 E2.3）。两个片段间最后的磷酸二酯键是由 **DNA 连接酶**来催化完成的，*E. coli* 中 DNA 聚合酶利用协同因子 NAD^+ 作为一个特别的能量来源。NAD^+ 被分解成烟酰胺单磷酸（NMN）和腺嘌呤单核苷酸（AMP），该 AMP 可通过 5′-5′连接共价结合到某一片段的 5′端。来自 NAD^+ 的焦磷酸键水解的能量以上述方式储存，并用于随后的该 5′端与相邻片段之间的 3′-OH 的连接，同时释放 AMP。体内的 DNA 聚合酶Ⅲ的全酶二聚体、引发体和 DNA 解旋酶以物理方式结合成一个大的称为**复制体**的复合体，该复制体以每秒 900bp 的速率合成 DNA。

终止与分离

两个复制叉在 *oriC* 约 180°的对面相遇。在这一区域内有几个终止子的位点，它们与 *tus* 基因产物即一种 DnaB 解旋酶的抑制剂结合，从而阻止复制叉的移动。这样即使 1 个复制叉由于某种原因而拖延了，它们仍会在终止处相遇。复制结束时，两个子链仍是相扣的，它们由拓扑异构酶Ⅳ（一种Ⅱ型拓扑异构酶）解联（参见 C4），结果随着膜的附着点移动而相互分离，分配到 2 个子细胞中（参见 D1）。

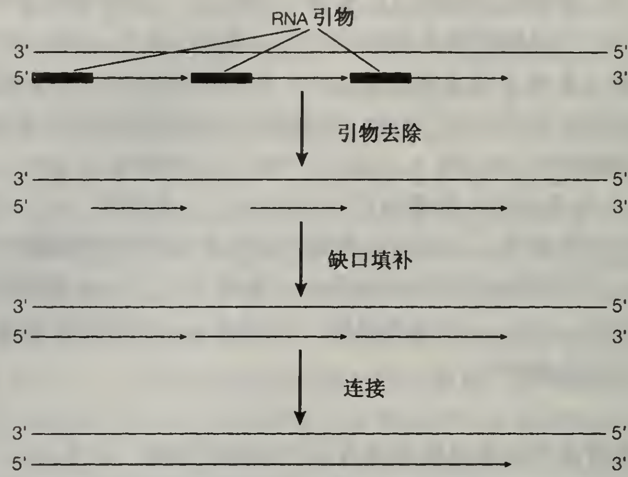


图 E2.3 后随链上 RNA 引物的去除以及随后的缺口填补与连接。DNA 聚合酶Ⅰ去除引物与填补缺口是同时进行的

E3　细胞周期

要　点

细胞周期	细胞周期包括 DNA 复制后的细胞分裂，即从一个母细胞产生两个子细胞。细胞周期是一个有序的过程，受细胞周期机制的控制。
细胞周期 4 个时期	细胞周期含有四个主要时期：G_1、S、G_2 和 M 期。S 期是 DNA 合成期；M 期（即有丝分裂）是细胞分裂期。这两个期被 G_1、G_2 这两个间隙期分隔开，M 期可以进一步划分为前期、中期和后期。G_1、S、G_2 共同组成细胞分裂的间期，细胞可由 G_1 期进入一个非增殖时期——G_0 期或称静止期。
检验点及其调控	细胞周期随细胞周围环境而调整，以免受损伤的细胞发生增殖。检验点是当外界条件不适合细胞分裂时可以中断细胞周期的一些阶段，主要检验点位于 G_1 和 G_2 的后期，在 G_1 期的 R 点，当缺乏促细胞分裂原时，细胞会脱离细胞周期进入非增殖的 G_0 期。
细胞周期蛋白和依赖于细胞周期蛋白的激酶	细胞周期是通过多种蛋白激酶复合物催化的蛋白磷酸化来实施调控的，这些复合物包含调控亚基（细胞周期蛋白）及催化亚基(依赖细胞周期蛋白的激酶，CDK)。不同的复合物调控细胞周期中的不同时期，它们的活性同时受到多方面因素的影响，如合成过程中的转录调控、抑制蛋白对其酶活的改变、蛋白水解对其造成的损伤。
E2F 和 RB 的调控	能否通过 G_1 期受控于转录因子 E2F 的激活，E2F 刺激细胞周期中所需基因的表达，在 G_1 期的早期 E2F 受到去磷酸化的 RB 结合的抑制。G_1 期中的细胞周期蛋白-CDK 复合物使 RB 磷酸化，从而释放 E2F 激活转录。
细胞周期的激活、抑制与癌症	GIP 和 INK4 蛋白通过抑制细胞周期蛋白-CDK 复合物的活性阻断细胞周期的进程，G_1 到 S 期的过渡受到原癌基因和抑癌蛋白的调控，B 细胞的癌变与细胞同期蛋白 D1 基因的过表达相关。RB 是 G_1 过渡到 S 期的一个关键调控因子，它和 INK4P16 都是癌症抑制蛋白。在 DNA 受到损伤时，CIP 蛋白 P21WAF1/CIP1 被抑癌蛋白 P53 所激活。

相关主题		
	真核生物的染色体结构（D3）	转录调控举例（N2）
	真核生物的 DNA 复制（E4）	癌基因的分类（S2）
	RNA 聚合酶Ⅱ基因：启动子与增强子（M4）	肿瘤抑制基因（S3）
		凋亡（S4）

细胞周期

细胞分裂的两个主要阶段是细胞 DNA 的复制和细胞分裂成两个子细胞，子细胞进一步分裂形成下一代细胞，细胞分裂过程的不断重复，也可看作是伴随着细胞分裂 DNA 复制的不断重复，这就是细胞周期。

细胞分裂是一个有序的和受到调控的过程。如果细胞没有先复制 DNA 就发生分裂或是在细胞分裂之前 DNA 发生了重复的复制，这将都是灾难性的。所以可以把细胞增殖看作一个循环过程，受到细胞周期机制的协调以保证细胞周期的正确顺序。

细胞周期 4 个时期

DNA 复制和细胞分裂发生在明显的且很有规律的时间间隔内，因此可将细胞周期模式化地分为四个时期（图 E3.1）

G_1

最长的一个时期（又叫**间隙期**），在此期间细胞为复制作准备

S

又称 **DNA 合成期**，DNA 进行复制，形成每条染色体的完整拷贝（参见 E4）

G_2

一个较短的间隙期，位于 S 期后、有丝分裂之前

M

有丝分裂，随着细胞分裂，染色体对等地分配进两个子细胞有丝分裂期可进一步分为**前期**，染色体发生凝集；**中期**，此间姐妹染色单体附着于着丝粒上（参见 D3）并在细胞中部相互成一条直线；**后期**，在此过程中姐妹染色单体分离并移向细胞两极最终形成两个子细胞。

G_1、S 和 G_2 共同组成**间期**，有丝分裂之后，增殖的细胞将进入下一个细胞周期的 G_1 期，细胞也可在有丝分裂之后脱离细胞周期而进入一个非增殖的休眠状态，G_0，也可称**沉默期**（quiescence）。

检验点及其调控

细胞分裂周期的起始需要胞外生长因子——**促细胞分裂原**。若缺少它的存在，细胞将会离开 G_1 期并进入 G_0 休眠期（图 E3.1）。在 G_1 期，细胞会对其环境进行评估，然后决定是否进入另一个分裂周期。G_1 期中的

这一点叫做**限制点**（或 **R 点**）。若细胞在到达 R 点之前缺乏促细胞分裂原，它将重进 G_0 期，而无法进行细胞分裂。而那些通过 R 点之后才开始缺乏促细胞分裂原的细胞会在进入 G_0 期之前实现细胞分裂完成细胞周期。在大多数细胞类型中，R 点发生在有丝分裂后的数小时。R 点无疑对了解进入细胞分裂周期的细胞定型是至关重要。在 G_1 期的有丝分裂和 R 点的期间，多种信号同时发生并互相作用最终决定细胞的命运。

像 R 点一样，细胞周期中可以终止细胞分裂或周期进程的那些点叫做**检验点**。检验点在间隙期发生作用，以保证细胞有能力进行又一轮的 DNA 复制（在 G_1 期的 R 点）及在细胞分裂之前成功完成 DNA 的复制（G_2 期检验点）。

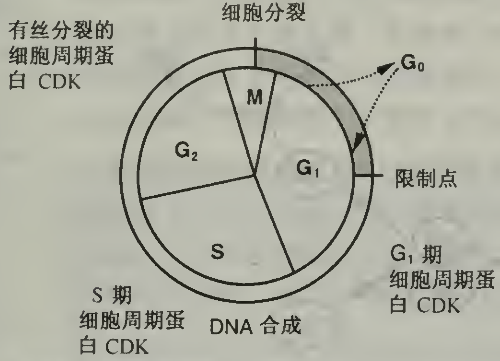


图 E3.1 细胞周期相图示。各期中重要的调控复合物用斜体表示

细胞周期蛋白和依赖于细胞周期蛋白的激酶

蛋白磷酸化是调控细胞周期进程的一个主要机制，是通过一个由调节亚基和催化亚基组成的蛋白激酶来实现的。调节亚基叫做**细胞周期蛋白**，催化亚基叫做**依赖细胞周期蛋白的激酶**（CDK）。CDK 只有与细胞周期蛋白发生结合，CDK 才有催化活性。每个 CDK 可与一种以上细胞周期蛋白结合。存在于一个特定的 CDK-细胞周期蛋白复合物中的 CDK 和细胞周期蛋白共同决定将被磷酸化的目标蛋白。

细胞周期蛋白-CDK 复合物有三种不同的种类，分别与细胞周期的 G_1、S 和 M 期有关。**G_1 CDK 复合物**通过激活的转录因子来保证细胞进入 S 期。为表达 DNA 合成时所需的酶和编码 S 期 CDK 复合物的基因。**S 期的 CDK 复合物**激发有序 DNA 合成的起始，该机制将确保每一染色体仅复制一次。**有丝分裂期的 CDK 复合物**诱导染色体凝集并有序地分配进两个子细胞。

CDK 复合物的活性受三条途径的调控：

(i) 对 CDK 复合物亚基转录的控制。

(ii) 降低 CDK 复合物的活性的抑制子，例如有丝分裂期的 CDK 复合物在 S 和 G_2 期合成，但直至 DNA 完成合成后它们才有活性。

(iii) 在不再需要 CDK 复合物的细胞周期中的特定时期 CDK 复合物的有

次序降解。

E2F 和 RB 的调控

细胞周期中由 G_1 进入 S 期部分地由基因转录激活（有时是抑制）来调控，而在随后的细胞周期中似乎主要被转录后机制所调控。通过 G_1 期中 R 点主要依靠对 **E2F** 这一转录因子的激活。E2F 刺激一些基因的转录与表达，这些基因编码为 DNA 复制、脱氧核糖核苷酸合成所需蛋白以及细胞周期蛋白和在后来的细胞周期中所需要的 CDK 蛋白。E2F 的活性受到与其结合的蛋白 **RB**（**成视网膜细胞瘤抑制蛋白**，参见 S3）及其相关蛋白的抑制。当 RB 处于去磷酸化状态时，E2F 活性受到抑制。在 G_1 中后期，细胞周期蛋白-CDK 复合物使 RB 磷酸化从而释放 E2F 进而激活转录(图 E3.2)。

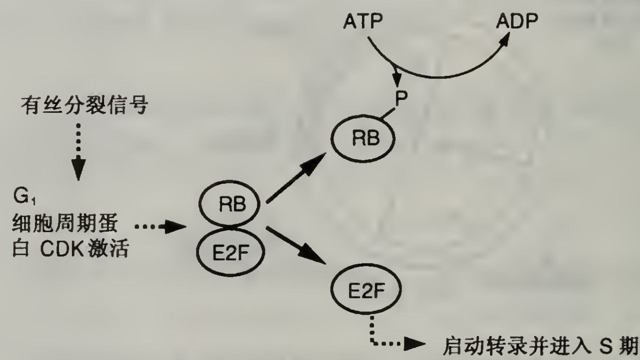


图 E3.2 E2F 受 RB 和 G_1 细胞周期蛋白-CDK 复合物调控的示意图

细胞周期的激活、抑制和癌症

小的抑制蛋白可以通过抑制细胞周期蛋白-CDK 复合物的活性延迟细胞周期的进程。这些抑制因子可分为两类：CIP 蛋白和 INK4 蛋白。例如在骨骼肌肉细胞分化时通过激活 CIP 蛋白基因中的一个（编码 P21WAF1/CIP1 蛋白），使分化细胞从细胞周期中脱离出来，这是由主调控因子 MyoD（参见 N2）所调控的。

细胞周期与癌症之间有着根本的联系，可以原癌基因（参见 S2）和抑癌基因（参见 S3）调控细胞由 G_1 进入 S 期的实验证据得到证明。G_1 细胞周期蛋白的一种，细胞周期蛋白 D1 在产生 B 细胞免疫球蛋白的肿瘤细胞中经常过量表达，这是因为染色体的移位使得免疫球蛋白基因增强了与靠近细胞周期蛋白 D1 基因的染色体某位点的结合。INK4 P16 蛋白拥有癌症抑制子的所有特征，重要的是，在人类癌症中两个重要的抑癌基因的产物：RB 和 P53（参见 S3）均与细胞周期调控密切相关，RB 参与调控 E2F 的活性；而 P53 相关于 DNA 损伤的细胞周期抑制。当 DNA 受到损伤时，P53 诱导 P21 WAF1/CIP1 基因的转录激活。

E4 真核生物的 DNA 复制

要 点

实验系统	较小的动物病毒如猿猴病毒 40（SV40）是研究延长的很好模型，但并不适合研究起始。仅有 400 个复制子的酵母比含有 50 000～100 000 个复制子的典型哺乳动物细胞要简单得多，非洲爪蟾（*Xenopus*）的卵提取物可有效地复制外加 DNA。
起始点与起始	在 S 期的不同时刻大约有 20～50 个起始点被激活。单细胞酵母的众多起始点已被克隆，且都有一个简单的 11bp 长的保守序列，该序列用以结合起始点识别复合体。一个特定因子可确保每个起始点在每个细胞周期中只被用 1 次。
复制叉	染色质解组装以复制 DNA 减慢真核生物复制叉的移动，并且在后随链上产生小片段。引发酶及 DNA 聚合酶 α 的聚合酶活性可起始前导链与后随链的复制，然后 DNA 聚合酶 δ 负责前导链和后随链的延伸，而 DNA 聚合酶 ε 完成后随链的复制。
核基质	由不溶性蛋白纤维构成的蛋白质支架在空间上控制复制。复制工厂被固定在核基质中，DNA 可穿过其中。
端粒的复制	端粒 DNA 由简单重复序列的多拷贝构成，3′端突出于另一条链的 5′端。端粒 DNA 由端粒酶催化复制，该酶携有用作重复片段模板的小 RNA 分子。端粒酶在体细胞中处于抑制状态，但在许多癌细胞中处于激活状态。
相关主题	原核与真核生物的染色体结构（D） DNA 复制概述（E1） 细胞周期（E3） 细菌的 DNA 复制（E2） DNA 修复（F3） DNA 病毒（R3）

实验系统 因为染色质结构的复杂性（参见 D2），真核生物复制叉以每秒约 50bp 的速度推移。以这样的速度，用 2 个复制叉来复制 10^5kb 这样大小的典型的哺乳动物的染色体 DNA 约需 30 天，因此需要多个复制子。一个典型的哺乳动物细胞中有 50 000～100 000 个复制子。同时研究这么多的复制子是不可思议的。所幸的是酵母（*Sacharomyces cervisiae*）的基因组较小（16 条染色体中含 14 000kb DNA），仅有 40 个复制子，目前全部的基因组序列都已公开。再简单一点的是病毒基因组如猿猴病毒 40（SV40）（参见 R3）。SV40 的 DNA 是一个 5kb 的环状双链，进入细胞核后可形成核小体。SV40 为研究哺乳动物复制叉提供了一个极好的模型。

另一个很有用的系统是从非洲爪蟾（*Xenopus laevis*）的卵中提出的无细胞提取物。该提取物中含有高浓度的复制蛋白（足以支持体内细胞的多次翻番），所以能广泛地用于外加 DNA 或整个细胞核的复制。

起始点与起始

在真核生物中，大约 20～50 个复制子（**串联排列**）**成簇**在 S 期的特定时间同时开始复制。S 期中较早复制的主要是**常染色质**（其中包括有转录活性的 DNA），较晚被激活的主要是**异染色质**内的部分（参见 D3），而着丝粒 DNA 及端粒 DNA 最后被复制。这种方式反映了不同的染色质结构与起始因子结合的难易程度。单细胞酵母的众多复制起始点已被克隆进原核生物的质粒。由于这些复制起始点序列允许质粒在酵母（真核生物）中复制，被称为**自动复制序列**（ARS）。可支持复制的最短 DNA 长度只有 11bp，其保守序列为［A/T］TTTAT［A/G］TTT［A/T］。但要达到最佳效果，需要这一序列的多个拷贝。这一序列被**起始点识别复合体**（**ORC**）所结合，被 CDK 激活后可引导 DNA 解链以进行复制。哺乳动物细胞的特定起始点序列还未被分离出来，研究者们认为每个复制子可能会在**起始区域**内的任意位置起始，而该区域可能有几个 kb 长，并且有可能是散在重复 DNA 片段的一部分（参见 D4）。同原核生物相比，真核生物的复制子在每个细胞周期仅起始一次复制。起始所必须的、并在作用后失活的特定蛋白（**特许因子**）只有在有丝分裂期核膜解体时才能进入核内，这样就防止了复制完成前的再次起始。

复制叉

复制前，DNA 必须在复制叉处从核小体上解离下来（图 E4.1），这一过程使得复制叉的移动速度减小到每秒 50bp。复制叉通过后，新的核小体由原来的和新合成的组蛋白重新组装起来。同原核生物一样，解开双链需要一种或多种 DNA 解旋酶，以及单链结合蛋白即**复制蛋白 A**（**RP-A**）。延伸过程涉及 3 种不同的 DNA 聚合酶。前导链及后随链的每个片段的复制都是利用 **DNA 聚合酶 α** 中的引发酶活性从 RNA 引物起始的。该聚合酶继续延伸 DNA 链，但很快在前导链上被 **DNA 聚合酶 δ** 代替，也许在后随链上也是如此。**DNA 聚合酶 ε** 的作用不很清楚，可能是在引物被移去后，DNA 聚合酶 ε 在后随链上完成 DNA 合成。这两种酶都兼有校对的活性。功能等同于 *E. coli* DNA 聚合酶Ⅲ全酶的 β 亚基的**增殖细胞核抗原**（proliferating cell nuclear antigen，PCNA）授予了 DNA 聚合酶 δ 合成长链 DNA 的能力。真核细胞后随链的短小片段的长度（如在 SV40 中是 135bp）反应了复制叉移动时解开每个核小体时缠绕在其上的 DNA 长度（图 E4.1）。在 S 期除了 DNA 的量得到加倍外，组蛋白的量也倍增了。

核基质

核基质是由不溶性蛋白纤维组成的支架，作为核中各种反应（包括 DNA 复制）进行时的组织架构行使功能（参见 D3），包括基因簇中所有复制子的复制叉上相关 DNA 以及所有蛋白质的巨大**复制工厂**被固定在基

质中，随 DNA 复制的进行，DNA 将穿过其中。通过脉冲标记即用胸腺嘧啶的类似物：**溴脱氧尿苷**（bromodeoxyuridine，BUdR）标记正在复制的DNA，可以在显微镜下观察到这些复制工厂。再通过用可识别 BUdR 的抗体的免疫荧光法可观察到被标记的 DNA（图 E4.2）。

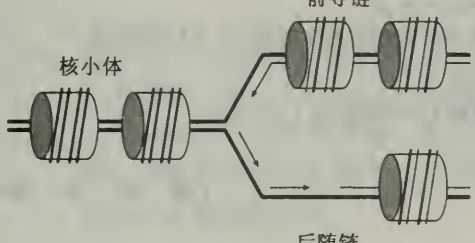


图 E4.1　真核生物复制叉附近的核小体

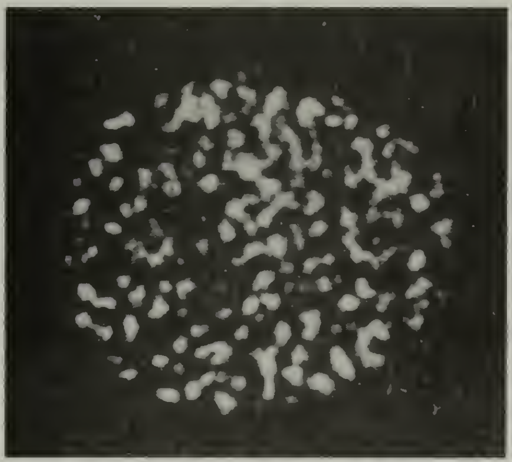


图 E4.2　真核细胞核中的复制工厂。经 Academic Press 的许可引自 H. Nakamura 等，1986，Exp. Cell Res. 165，291～297

端粒的复制

半不连续复制机制不能完成线性染色体末端的复制，因为没有 DNA 可用来取代从后随链的 5′端切掉的 RNA 引物作为模板来延伸相应的 DNA。这样遗传信息很可能会从 DNA 上丢失。为克服这一缺点，真核生物染色体的末端（端粒，参见 D3）采取了由数百个简单重复的序列构成的结构（例如在人类中是 TTAGGG），这些序列不含遗传信息，并且 3′端突出于另一条链的 5′端（图 E4.3）。端粒酶含有一个约 150 核苷酸长的 RNA 分子，其部分序列与上述重复序列互补。该 RNA 分子作为模板，通过反复延伸（聚合作用）与易位，反复地将重复片段加到突出的 3′端上，而互补链则像一般的后随链那样复制，最终留下 3′突出端。有趣的是端

粒酶的活性在多细胞生物体细胞中被抑制，结果是每一细胞子代中染色体的长度都渐渐缩短。当该长度缩短到含遗传信息的 DNA 碱基时，细胞就会衰老然后死亡。这种现象在细胞老化中也许是很重要的，而许多癌细胞的无限增殖能力就是与端粒酶活性的再次激活相关的。

3'-AATCCCAATCCC-5'
5'-TTAGGGTTAGGG (TTAGGG)$_n$ TTAGGG-3'

图 E4.3 人染色体端粒 DNA 的序列（n = 几百）

（杨 涛 译 刘进元 校）

F1 诱变

要 点

突 变

突变是 DNA 碱基序列水平上的永久性变化且可遗传。点突变可以是转换（如 G·C→A·T）或颠换（如 G·C→T·A）。缺失与插入涉及碱基的丢失或增添，并且可以引起遗传密码可读框的移码。沉默突变没有表型效应，而错义突变和无义突变会改变所编码蛋白的氨基酸序列。

复制忠实性

DNA 复制的高精度（10^{10}个碱基中约有 1 个错配）依赖于模板链和进入核苷酸在 DNA 聚合酶的作用位点的正确配对、3′→5′外切核酸酶对掺入碱基的校正以及错配修复三方面的精确配合。

物理诱变剂

离子化的（如 X 射线和 γ 射线）和非离子化的（如紫外线）辐射会产生各种各样的 DNA 损伤。嘧啶二聚体是紫外线辐射的最常见产物。

化学诱变剂

碱基类似物可以在 DNA 复制过程中以错配方式导致突变。亚硝酸可使胞嘧啶和腺嘌呤脱氨。烷化剂和芳基化试剂可产生一系列中断转录与复制的加合物，从而直接引发突变或更常见的是间接导致突变。大部分化学诱变剂是致癌的。

直接诱变

如果一种碱基类似物或配对特性与亲本链碱基不同的修饰碱基在复制叉通过之前未被 DNA 修复机制去除，结果会引入 1 个错配碱基，而第 2 轮复制会将这一突变在 DNA 中永久地固定下来。

间接诱变

在复制叉通过之前，DNA 中的大部分损伤都会被直接地无差错恢复或切除修复机制所修复。如果未能被修复，将与特定 DNA 聚合酶有关的一种转移损伤 DNA 合成的易错形式就会发生，结果是一个或多个错配碱基被掺入在损伤部分的对应位点。

相关主题

核酸结构（C1） DNA 损伤（F2）
DNA 复制（E） DNA 修复（F3）

突变

突变是 DNA 碱基序列水平上的永久性的、可遗传的改变。突变产生于 DNA 复制或减数分裂重组过程中的自发性错误（参见 F4），或者是由于物理或化学试剂损伤 DNA 所致。最简单的突变是**点突变**，即一个单一碱基的改变。点突变可以是**转换**（嘌呤与嘌呤之间，嘧啶与嘧啶之间互换）或**颠换**（嘌呤与嘧啶之间发生互换）。点突变的表型效应是多样化的，

如果它发生在 DNA 的非编码区、非调节区或密码子的第 3 个碱基位置，那么它就不会影响掺进蛋白质中的氨基酸（参见 P1），则该突变是**沉默突变**；如果它改变了基因产物的 1 个氨基酸，则是**错义突变**。错义突变的效应可以是无作用的，也可以是致死的，这取决于被影响的氨基酸。形成新的终止密码子的突变是无义突变，结果会产生截短了的蛋白质产物。插入或缺失涉及一个或多个碱基的增加或丢失，这会引起基因的**移码突变**。移码突变的结果是所翻译出的蛋白质序列从突变点起至 C 末端都完全被改变了。影响细胞生长或细胞死亡过程的突变可以引发癌变（参见 S)。群体中许多沉默及非致死突变的积累会产生**遗传多态性**，即在“正常” DNA 及其蛋白质序列中可接受的变异（参见 D4)。

复制忠实性

DNA 复制的错误率要比转录的低得多，这是维持上一代与下一代间遗传信息的准确含义的需要。例如，*E. coli* 在复制中的**自发突变**率大约是每掺入 10^{10}碱基会发生一个错配。这主要是由于可改变其碱基配对性质的互变异构体稀有碱基的存在（参见 C2)。可通过多种机制将错误率降低到最小程度。只有当进入活性位点的核苷酸与模板核苷酸满足正确的 Watson-Crick 碱基配对原则时，DNA 聚合酶才能将其掺入进来。聚合酶的 3′→5′外切酶活性可检测出偶发错误（图 F1.1a)（参见 E2)。这将在下一个核苷酸掺入前从 3′端去掉错配核苷酸，从而保证正确碱基的插入 (图 F1.1b)。**校对外切核酸酶**必须能区分正确的与错误的碱基对，才能正常工作。DNA 后随链的新起始片段的 5′端“未锚定”的几个碱基对的较大流动性（参见 E1）意味着它们决不会正确配对，故不能被校对。这是因为头几个核苷酸是 RNA，它们后来被识别为低忠实性碱基，结果被来自邻近片段延伸（并校对）的 DNA 替代。逃脱校对的错误可被错配修复机制所纠正（参见 F3)。

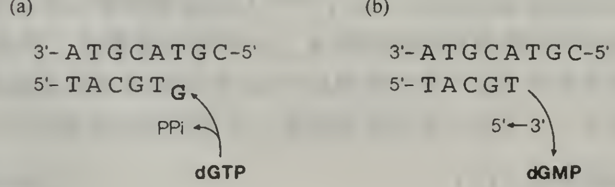


图 F1.1　新合成 DNA 的校对。例如，一个“G”碱基错误地与模板“T”结合 (a)，它将被 3′→5′校对外切酶所切除 (b)

物理诱变剂

高能离子化辐射（如 X 射线，γ 射线）的吸收会引起目标分子失去电子，这些电子会引起 DNA 发生广泛的化学变异，包括断链、碱基及五碳糖的损伤。非离子化辐射会引起在目标分子内的分子振动或促进电子进入较高能级，导致形成新的化学键。引起 DNA 损伤的最重要的方式是紫外线，紫外线引起相邻嘧啶碱基产生**嘧啶二聚体**（参见 F2)。

化学诱变剂

碱基类似物是改变碱基配对特性的正常碱基衍生物，可诱发**直接诱变**。许多天然的、合成的有机和无机的化学试剂均可与 DNA 发生反应并改变其特性。**亚硝酸**使胞嘧啶脱氨变成尿嘧啶（图 F1.2），从而和腺嘌呤配对，引起随后复制中发生 G·C 向 A·T 转换。腺嘌呤脱氨变成鸟嘌呤类似物次黄嘌呤，导致 A·T 向 G·C 转换。**烷化剂**如**甲烷磺酸甲酯**（methylmethane sulfonate，MMS）、**乙基亚硝基尿**（ethylnitrosourea，ENU）及**芳基化剂**通常会产生须经修复才足以防止给转录及复制过程带来严重破坏的损伤（参见 F2）。细胞对于这些损伤的处理有可能会由于**间接诱变**而导致突变。嵌入剂（参见 C4）会产生插入突变和缺失突变。绝大部分化学诱变剂是**致癌剂**，诱发癌症。

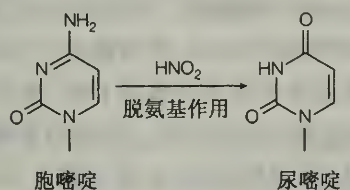


图 F1.2 亚硝酸使胞嘧啶脱氨生成尿嘧啶

直接诱变

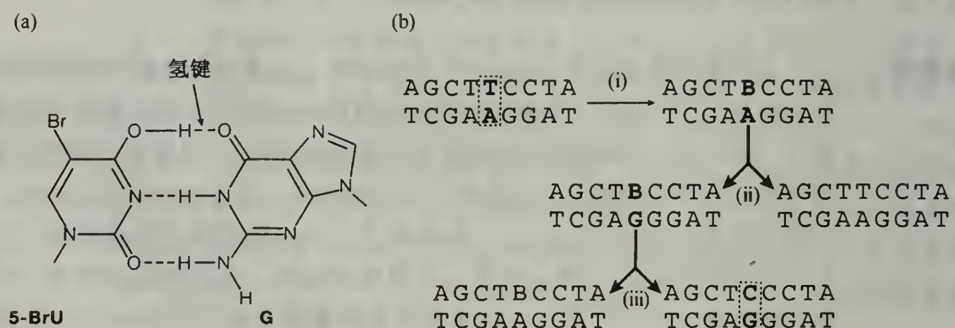
直接诱变是起因于 DNA 中存在稳定的、具改变了的配对特性的未修复的碱基。正是 DNA 的复制使得这种损伤（非永久性的）固定下来成为突变（永久性，可遗传的）。例如胸腺嘧啶类似物 5-溴尿嘧啶的酮基异构体预期会与腺嘌呤配对，也经常出现以烯醇式（这是由于溴的负电性）与鸟嘌呤配对（图 F1.3a）。一个复制回合之后，鸟嘌呤也许就被插入到与 5-溴尿嘧啶对应的位置。在第 2 个复制回合后，胞嘧啶就会被掺入到与该鸟嘌呤相对应的位置上，这样突变就被**固定**下来了（图 F1.3b）。最终结果是 A·T 向 G·C 的转换。8-氧-dGTP 是细胞内氧化产生的一个天然的 dGTP 的衍生物（参见 F2），可与 A 配对，当它结合到 DNA 中后，可导致 A·T 向 G·C 颠转。

间接诱变

化学和物理诱变剂导致损伤的绝大部分，在与复制叉相遇之前会成为一种或多种 DNA 修复机制作用的目标，从而恢复到无差错状态的 DNA 原初结构（参见 F3）。有时这是不可能的，比如当损伤就发生在移动着的复制叉之前，将产生 1 个不可"**重组修复**"的损伤（参见 F4）。这会导致在子链上产生一个致命的缺口，因此在这种情况下细胞不得不采取"**转移损伤 DNA 合成**"。这就包括正常的复制装置被一种具有特殊 DNA 聚合酶活性的装置所临时代替，而这种装置则在未修复损伤的对应位点上插入碱基，然后又恢复正常复制。因此损伤没有被转移或修复而是被容忍，从而保持复制 DNA 的完整性，在下一轮复制之前模板链可以被修复。

在一些情况下，转移损伤 DNA 合成有可能正确从而插入正确的碱

基，但被破坏的碱基往往失去了特殊的碱基配对性，所以一些损伤DNA聚合酶只是为了保证染色体的完整性在损伤对应的位点上插入错误的碱基，这就叫**间接诱变**。如果突变只发生在损伤的位点上，该突变就是“**定点的**”；如果基因组的其他位点也发生了突变，则该突变就是“**非定点的**”。在原核生物中，转移损伤DNA合成是对DNA损伤的**SOS反应**中的一部分。虽然它不是一个严格的修复机制，有时也称之为“**易错修复**”。SOS反应涉及许多能在DNA损伤后提高存活的相关基因的诱导。损伤诱导性RecA，UMUc，UmuD和DinB蛋白对于由间接诱变所产生的突变来讲是十分必要的。UmuC和UmuD的一种修饰形式（UmuD′）一起形成一个蛋白复合体（$UmuD'_2C$），最近被称为DNA聚合酶Ⅴ，可以在像“**脱嘌呤**”位点（参见F2）这样的“**非编码**”损伤对应的位点上插入一个碱基从而导致定点突变。另一种称为DNA聚合酶Ⅳ的转移损伤聚合酶DinB会诱导非定点突变。非定点突变是对有害的DNA损伤环境的故意反应，为了增加突变频率从而产生能更具生存的、抗性的菌株。在真核生物中，转移损伤DNA聚合酶包括易错型的DNA聚合酶ζ（zeta）和可能为无错型的DNA聚合酶η（eta）。



图F1.3 (a) 5-溴尿嘧啶以烯醇式与鸟嘌呤相配对；(b) 5-溴尿嘧啶（B）引发的直接诱变：(i) T的类似物B可很容易地掺进T的位置；(ii) 一次复制后G被掺入到“B”的相对位置；(iii) 2次复制后，C将被掺入进G相对应的位置

F2 DNA 损伤

要 点

DNA 损伤 外源化学试剂或射线对 DNA 的化学作用会导致其化学或物理结构发生变化。这些变化也许会阻断复制或转录，结果是致死性的；或者会通过直接或间接诱变产生突变。DNA 的化学不稳定性可产生自发性损伤如脱氨和脱嘌呤。

氧化性损伤 活性氧分子如超氧化物及羟自由基会产生种种损伤（如 8-氧化鸟嘌呤和 5-羟甲基尿嘧啶）。这样的损伤是自发产生的，而且包括 γ 射线在内的一些外源物质还会提高其发生率。

烷基化 亲电的烷化剂像甲烷磺酸甲酯（methylmethane sulfonate，MMS）和乙基亚硝基尿（ethylnitrosourea，ENU）能在许多位点上对核苷酸进行修饰。绝大多数损伤是间接诱变的，而 O^6-烷基鸟嘌呤（O^6-alkylguanine）是直接诱变的结果。

聚化加合物 大块损伤像嘧啶二聚体和烷化剂加合物会使得双螺旋扭曲变形，引起局部变性，结果是扰乱了 DNA 的正常功能。

相关主题 核酸结构（C1） DNA 修复（F3）
诱变（F1） 凋亡（S4）

DNA 损伤 损伤是 DNA 正常的化学或物理结构的改变。碱基杂环上的一些 N 和 C 原子及一些环外功能团（如碱基的酮基及氨基）是具有相当的化学活性的。许多外源物质如化学试剂及射线能引起这些位点的改变。碱基的化学特性的变化会导致不能配对或错配（如一个改变了的 A 可能会与 C 而不是与 T 配对）。如果这种损伤允许在 DNA 中保留下来，这一突变就会通过**直接诱变**或**间接诱变**而被固定下来（参见 F1）。或者这些化学改变也可能造成 DNA 的结构变形从而阻断复制或（和）转录，导致细胞死亡。因此 DNA 损伤可以是**致突变**和（或）**致死**。一些损伤是**自发性的**，是因为 DNA 内在的化学活性以及细胞中存在的正常活性化分子所致。例如胞嘧啶会自发地水解**脱氨**变成尿嘧啶（参见 F1，图 F2.2）。如果未被修复，产生的尿嘧啶会在接下来的复制中与腺嘌呤配对，从而产生点突变（参见 F1）。事实上 DNA 中以这种方式产生尿嘧啶很可能是为什么 DNA 含有胸腺嘧啶而不是尿嘧啶的原因。DNA 中发现的任何尿嘧啶均可被一种称为尿嘧啶 DNA 糖化酶所切除，并由胞嘧啶所替代（参见 F3）。5-甲基胞嘧啶，一个在 DNA 中少量存在的修饰碱基（参见 D3）会脱氨变成胸腺嘧

啶，即一个正常碱基，而检测这种变化是非常难的。

脱嘌呤作用是另一种在嘌呤 A 和 G 的 N－9 及脱氧核糖 C－1′之间**N－糖苷键**发生断裂的自发水解反应，DNA 因此失去了嘌呤碱基。DNA 的糖-磷酸骨架仍然是完整的，产生的**脱嘌呤位点**呈现"**非编码损伤**"，即该嘌呤碱基所编码的遗传信息丢失了。脱嘌呤作用在 37℃下以每小时丢失 10 000个嘌呤的速度在人体中发生。**脱嘧啶**也可能发生，但频率很低。

氧化性损伤

在所有需氧细胞中由于超氧化物、氢过氧化物及最重要的羟基自由基等活性氧（ROS）的存在，会在正常条件下会发生氧化损伤。这些自由基可在许多位点上攻击 DNA，产生一系列特性变化了的氧化产物，如 8－氧化鸟嘌呤，2－氧化腺嘌呤和 5－羟甲基尿嘧啶（图 F2.1）。电离辐射引起的水辐射分解所产生的羟基自由基会提高这些氧化产物的水平。

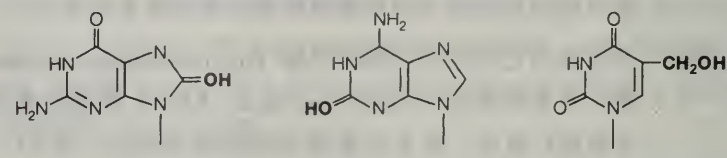


图 F2.1 活性氧作用造成在 DNA 上形成的氧化碱基的例子

烷基化

烷化剂是可将烷基（如甲基）加入到核酸上各种位点的亲电化学试剂，但其加入位点有别于正常甲基化酶的甲基化位点（参见 C1 及 G3）。常见的烷基化试剂有 **MMS** 和 **ENU**（图 F2.2a）。甲基化碱基的典型例子是 7－甲基鸟嘌呤、3－甲基腺嘌呤、3－甲基鸟嘌呤和 O^6－甲基鸟嘌呤（图 F2.2b）。这些损伤中的部分由于会在 DNA 复制及转录时干扰 DNA 解旋，因而可能是致死的。绝大部分是间接诱变损伤；而 O^6－甲基鸟嘌呤由于可在复制中与胸腺嘧啶配对，产生直接诱变损伤。

聚化加合物

紫外线照射可通过 DNA 链上相邻嘧啶每个碱基的 C5 双键和 C6 碳原子环化形成一个环丁烷环从而形成**环丁烷嘧啶二聚体**（图 F2.3a）。结果不能与其相对应的链进行碱基配对，导致 DNA **局部变性**，产生破坏复制与转录的大块损伤。另一种嘧啶二聚体 **6，4－光产物**，则是由一个嘧啶上的 C6 和其相邻碱基上的 C4 间生成键后形成的（见图 F2.3a 中环上的碳数字）。当煤焦油中的致癌物**苯并芘**在肝内由细胞色素 P－450 代谢后，其中的一种代谢物（一种二醇环氧化物）可与鸟嘌呤残基共价结合（图 F2.3b）。许多其他的**芳香族烷化剂**均能与 DNA 形成共价加合物。致癌物黄曲霉毒素 B_1（aflatoxir B_1）也可与 DNA 共价结合，导入大块加合物，从而在 DNA 复制过程中改变遗传信息。

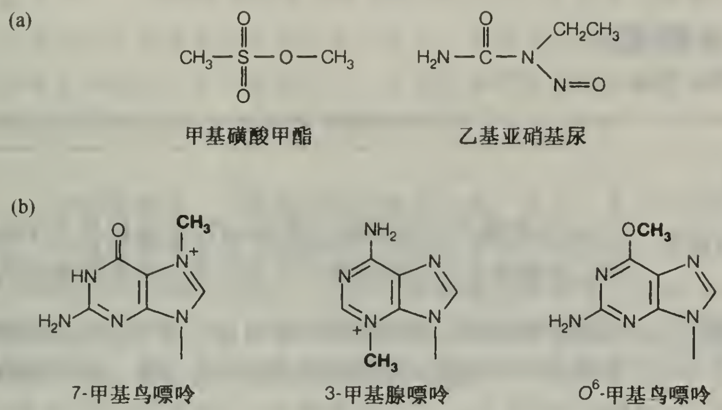


图 F2.2 烷化剂（a）和烷化碱基（b）的例子

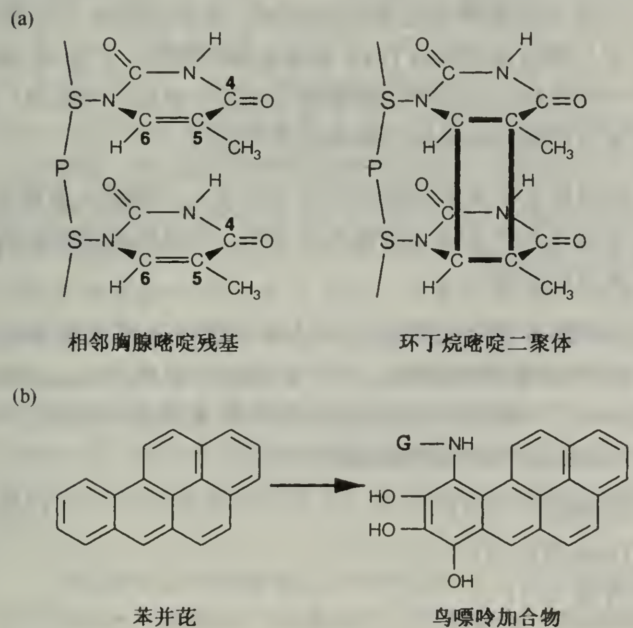


图 F2.3 相邻胸腺嘧啶残基形成环丁烷嘧啶二聚体（a），S＝糖，P＝磷酸，苯并芘二醇环氧化物的结合所形成的鸟嘌呤加合物（b）

F3　DNA 修复

要　点	
光复活	DNA 光解酶可切开嘧啶二聚体的环丁烷恢复其 DNA 的原初结构。光解酶含有可吸收蓝光为反应提供所需能量的色素分子。
烷基转移酶	烷基转移酶是一种可诱导的蛋白质，可特异地从鸟嘌呤的 O^6 位置移去一个烷基，将其转移到蛋白质自身，同时导致该蛋白质失活。
切除修复	在核苷酸切除修复中，内切核酸酶在损伤部位两边切出切口，去除损伤部分而留下一个缺口，该缺口由 DNA 聚合酶来填补，DNA 连接酶将催化最后的磷酸二酯键的形成。在碱基切除修复中，损伤被特定的 DNA 糖基化酶所切除，产生的 AP 位点被 AP 内切核酸酶以及外切核酸酶切除并扩展成一个缺口，接下来的修复过程就像核苷酸切除修复那样进行。
错配修复	逃过校对的复制错误在子链上存在一个错配。复制后 DNA 的半甲基化可将子链与母链区别开来，子链中的错配碱基可被切除修复机制去除。
遗传性的修复缺陷	切除修复基因或损伤性 DNA 聚合酶中的突变会导致不同形式的 DNA 修复酶缺乏病，又称着色性干皮病（xeroderma pigmentosum)，一种对太阳光敏感的有癌变倾向的紊乱，切除修复在 Cockayne 综合征中也是缺乏的。
相关主题	核酸结构（C1）　DNA 损伤（F2） DNA 复制（E）　凋亡（S4） 诱变（F1）

光复活　在可见光存在的情况下，**DNA 光解酶（光复活酶）**可将环丁烷嘧啶二聚体再分解为单体。这些酶含有可吸收蓝光并将能量转移到待切环丁烷环中的辅基（参见 B2）。*E. coli* 的光解酶含有 2 个色素分子，N^5，N^{10}-次甲基四氢叶酸和还原性的黄素腺嘌呤二核苷酸（FAD)。光复活对嘧啶二聚体是专一性的，是损伤被“直接修复”的一种例子，是无差错的。

烷基转移酶　另一无差错直接修复的实例构成了 DNA 对烷化剂的**适应性反应**。*E. coli* 中低剂量烷化剂会诱发这一反应并提高了机体对高剂量的诱变效应和致死的保护能力。能直接从突变的 O^6 -烷基鸟嘌呤（可与胸腺嘧啶配对，

参见 F2）上去除烷基的烷基转移酶可保护 DNA 免受这类诱变，奇怪的是烷基被转移到该蛋白本身并使之失活。这样每个烷基转移酶只能用一次。在哺乳细胞中也含有这种蛋白。致死性保护还涉及由**碱基切除修复**去除其他的烷化碱基的 DNA 糖基化酶的诱导。

切除修复

切除修复是一种普遍存在的修复机制，可在一系列的损伤中起修复作用，并且这种修复是无差错的。切除修复有 2 种形式，即**核苷酸切除修复**（nuclotide excision repair，NER）和**碱基切除修复**（base excision repair，BER)。在 NER 中，内切核酸酶在损伤部位两边各切除精确数目的碱基（图 F3.1a，第 1 步)，然后包含损伤的寡聚核苷酸被切除并留下一个缺口（图 F3.1a，第 2 步)。举例来看，*E. coli* 中的 UvrABC 内切核酸酶识别双螺旋中发生变形的部位，切除嘧啶二聚体和其他大块损伤。在 BER 中，相当专一的 **DNA 糖基化酶**识别修饰碱基，切除修饰碱基与糖基之间的 *N*－糖苷键（参见 C1)，留下一个**脱嘌呤**或**脱嘧啶**（AP）位点（图 F3.1b，第 1a 步)。自发碱基丢失也可产生 AP 位点（参见 F2)。**AP 内切核酸酶**在该位点切开 DNA，其外切酶活性可继续切出 1 个缺口（图 F3.1b，第 1b 及 2 步)。NER 中的缺口通常更大些，而在 BER 中可小至 1 个核苷酸。从这一点来看，两种形式的切除修复本质上是相同的。*E. coli* 中缺口可由 DNA 聚合酶 Ⅰ 填补（图 F3.1，第 3 步)，最后的磷酸二酯键的形成由 DNA 连接酶完成（图 F3.1，第 4 步)，很像 DNA 复制中后随链片段复制过程中的最后一步（参见 E1)。在真核生物中，BER 中的缺口主要由 **DNA 聚合酶 β** 来填补，而 NER 中较长的缺口主要由 DNA 聚合酶 δ 或 ε 来填补。在真核生物 NER 中，DNA 损伤的识别和切除是一种至少与 18 种多肽因子相关的复杂过程，这其中包括转录因子 TFⅡH（参见 M5)。切除修复是与转录相偶联的，因而被转录（一般是活性的）DNA 修复比未被转录 DNA 区域修复得更快，这有助于限制缺陷基因产物的生成。

错配修复

错配修复是切除修复的一种特定形式，用于修复在复制中错配并漏过校正检验的任何碱基（参见 F1)。复制错配中的错配碱基存在于子代链中。因此该系统必须在复制叉通过之后有一种能识别亲本链与子代链的方法，以保证只从子代链中去除错配碱基。在原核生物中，两条链上 GATC 序列中的某些腺嘌呤残基正常是甲基化的（参见 C1)，子代链的甲基化在复制完成几分钟后随即进行。这样新复制的 DNA 是**半甲基化的**，即亲本链是甲基化的而子代链未甲基化，所以很容易区分它们。错配的碱基对（如 GT 或 CA）被由 MutS 和 MutL 蛋白组合的复合体识别并与之结合，该复合体随后再与 MutH 内切核酸酶相结合，后者在子代链 GATC 附近的位点上特异性的产生缺刻。这个缺刻启动了对含有错误碱基区域的切除修复。真核生物中区分的机制还不清楚，但显然错配修复对保持 DNA 复制中的总体错误速率即自发突变率相当重要，比如**遗传性非息肉结肠癌**就

是由于一种错配修复酶突变丢失引起的。错配修复还可以修复真核生物**减数分裂重组**中由于序列错误对排而产生的错误。

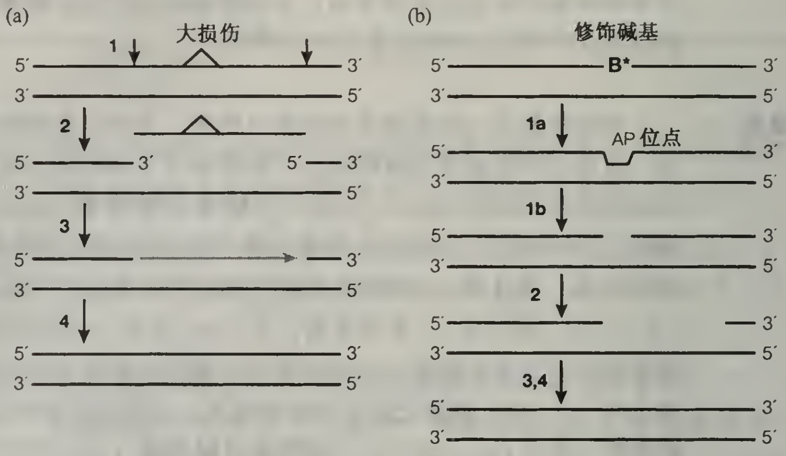


图 F3.1　(a) 核苷酸切除修复；(b) 碱基切除修复。标号步骤在文中有说明

遗传性修复缺陷

着色性干皮病（xeroderma pigmentosum，XP）是一种常染色体隐性紊乱，在表型上表现为对阳光极度敏感，极易产生皮肤癌。XP 患者缺乏包括由紫外线引起的大块 DNA 损伤的核苷酸切除功能，至少有 7 种不同基因的缺陷可导致 XP，表现出哺乳动物细胞中切除修复的复杂性。**着色性干皮病的变异型**（XP-V）在临床上与经典 XP 很相似，但是 XP-V 患者的细胞能执行正常的 NER。在这种病中，缺陷位于编码转移损伤 DNA 聚合酶 η 的基因中，这种酶在正常情况下可以在环丁烷胸腺嘧啶二聚体（参见 F1 与 F2）中的胸腺嘧啶残基相应的位点上无错性的插入 dAMP。因此，XP-V 细胞就更多的依赖于转移损伤 DNA 合成的另一种易错型模式以保证 DNA 在受辐射损伤后的完整性。**Cockayne 综合征**患者也对阳光敏感，并在转录偶联的切除修复中有缺陷，但没有诱发癌症的倾向。

F4 重组

要 点

同源重组	在真核生物的有丝分裂过程中，广泛存在着两条 DNA 分子之间同源区域的交换。在原核生物中依靠 *recA* 的重组产生一个四条链的 Holliday 中间体，后者可通过两条途径拆分。同源重组复制可使未修复损伤的 DNA 恢复原来的完整性。
位点特异性重组	发生在非同源区域中特定位点的 DNA 交换要依靠 *recA* 实现。把 λ 噬菌体整合到 *E. coli* 的基因组中就是通过这种重组，该过程需要两者都含有一个特定的 15bp 的序列以及一些特定的蛋白整合因子。位点特异性重组同时也解释了在动物体内抗体多样性的形成原因。
转座作用	可转座的 DNA 元件可以使自身插入到基因组的任何位置。所有的转座子都编码一个转座酶用以催化插入。反转录转座子通过一个 RNA 中间体复制并且与反转录病毒有关。

相关主题 细菌 DNA 复制（E2） DNA 修复（F3）
诱变（F1） RNA 病毒（R4）

同源重组 也称**一般重组**。该过程涉及两个 DNA 分子间同源区域的交换。对于双倍体真核生物，同源重组经常发生在**减数分裂过程**，在其中期Ⅰ同源复制的染色单体平行排列，非姐妹染色单体通过**交换**相对应的区域。有丝分裂结束后，产生的单倍体配子就会包含来源父母本两方面的遗传信息，这样就可以保证个体继承多方的基因信息。单倍体细菌也进行重组，如发生在部分已复制 DNA 间或发生在染色体 DNA 与获取的外源 DNA（如质粒或噬菌体）之间。在 *E. coli* 中二个同源 DNA **双螺旋**排列在一起（如图 F4.1a），通过核酸酶和 **RecBCD** 蛋白复合体的作用在一对**姐妹**染色单体（序列相同的单体）上产生切口，切口具体发生在称为 *chi*（GCTGGTGG）的特定序列上，该序列大约每 4kb 就有一个（图 F4.1b）。含有 5′端切口的单链 DNA 被 RecA 蛋白包裹形成 RecA-ssDNA 细丝，这些细丝相互交换并寻找相对的 DNA 双螺旋上的相应序列（**侵入**，如图 F4.1c），这之后切口被封好并形成四分支的 **Holliday** 结构（图 F4.1d）。这个结构是动态的，交叉点左右任何两个方面移动相当一段距离（**分支迁移**，如图 F4.1e），这个 Holliday 中间体能以两种方式中的一种拆分为两个 DNA 双螺旋。如果两条侵入链被切断，那么所产生的**重组体**除了中间的交换部分

之外与原来的分子都是一样的（如图 F4.1f）。如果是非侵入链被断开，则产生的重组体一半来自父本，另一半则来自母本，形成的两个亲本的**杂合双螺旋**（图 F4.1g）。同源重组对 DNA 修复也很重要，当复制叉遇到一个未修复的非编码的损伤，它能“跳过”DNA 损伤区域并继续复制，留下一个**子链缺口**，通过重组将亲本姐妹链上相应区域替代并填平缺口。由此产生的亲本姐妹链上的缺口很容易被填平，因为它的对面没有损伤，原来损伤部分可通过正常的切除去掉。这个机制也被称作**复制后修复**。

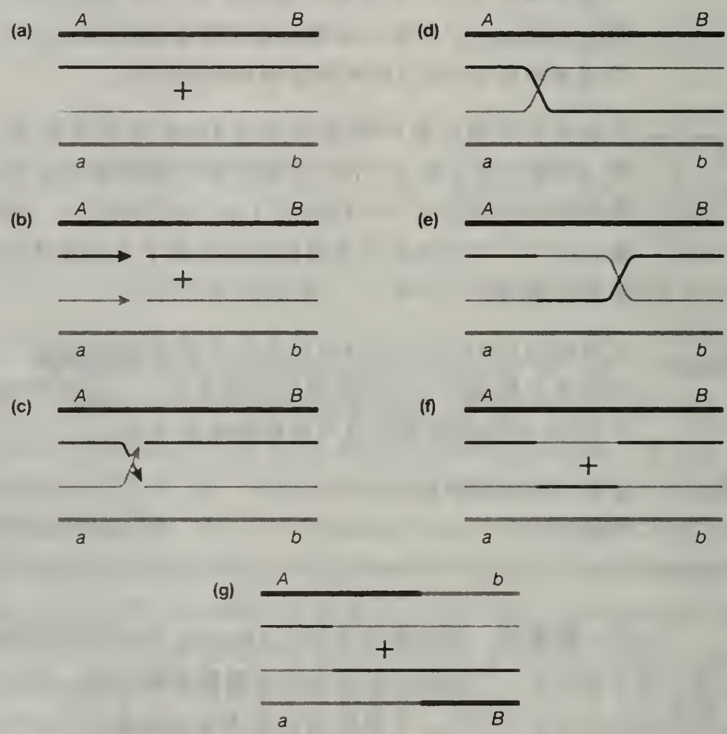


图 F4.1 2 个 DNA 双螺旋分子间同源重组的 Holliday 模型

位点特异性重组

位点特异性重组指非同源 DNA 的特异片段之间的交换，由能识别特异 DNA 序列的蛋白质所介导，并不需要 RecA 或单链 DNA。λ 噬菌体（参见 R2）可将自身基因组插入 *E. coli* 染色体的特定位点，两者 DNA 上的连接位点处有 15bp 的序列是相同的。λ 噬菌体编码的**整合酶**（Int）可在这些序列上形成具 7 个 bp 的突出端的交错切口，并与细菌编码的**整合宿主因子**（IHF）一起促成这些位点与 λ－DNA 间的重组结果和 λ－DNA 进入宿主染色体。这种方式形成的 λ 噬菌体称为**原噬菌体**，可随细胞分裂而稳定遗传，直至 λ 噬菌体所编码的**切除酶**被激活，整个过程再颠倒过来。在真核生物中，位点特异性重组可以解释抗体多样性的形成。免疫球蛋白由不同类型的两条重链和轻链组成，都包含不变的和可变的氨基酸序列。其中可变区域在生殖细胞中由 3 个基因段所编码：*V*、*D* 和 *J*。

有 250 个 *V*，15 个 *D* 和 5 个 *J* 基因用以编码重链，250 个 *V* 和 4 个 *J* 编码轻链。在抗体产生细胞的分化过程中，这些基因段之间的重组能产生大量（>10^8）的不同重链和轻链基因序列，进而导致抗体的种类多样化。每个 *V*、*D* 和 *J* 基因段都与短的识别序列相连用以形成重组蛋白。*V* 基因跟有一个 39bp 的序列，*D* 基因与 2 个 28bp 的序列相连，而 *J* 基因之前有一个 39bp 片段。由于重组只能发生在 39bp 和 28bp 序列之间，这就保证了只能形成 *VDJ* 体。

转座作用

转座子或**可转座元件**是一些较短的 DNA 序列，可以转移进细胞基因组的几乎任何位置。转座作用又被称为**异常重组**，因为它不需要序列间具有同源性，也不是位点特异性的，结果它的效率较低。最为简单的转座子是 *E. coli* 中的 **IS 元件**或**插入序列**，它在基因组中可能有几个拷贝。这些拷贝长约 1～2kb，包含一个**转座酶**基因，两侧为短的（～20bp）**末端反向重复序列**（相同的序列但方向相反）。转座酶在染色体 DNA 上形成交错切口，在复制过程中转座子的一个拷贝插入到目标位点中（图 F4.2），缺口在 DNA 聚合酶 Ⅰ 和 DNA 连接酶的作用下填平，产生目标位点的一个副本并形成一个新的**同向重复序列**。被转座子插入的基因一般会失活，在两个转座子拷贝之间的基因可以通过它们之间的重组而被切除。宿主 DNA 序列的颠倒或其他方式的重排也可能发生。**转座子诱变**是产生突变体的一个有效方法。除了转座酶，**Tn** 转座子系列还携带其他基因，其中包括 **β-内酰胺酶**，可使生物体抗青霉素。一些抗性基因通过转座作用进入质粒，导致抗生素抗性在细菌中传播，而质粒在细菌中很容易复制。转座子也具有穿越原核种间障碍的能力。

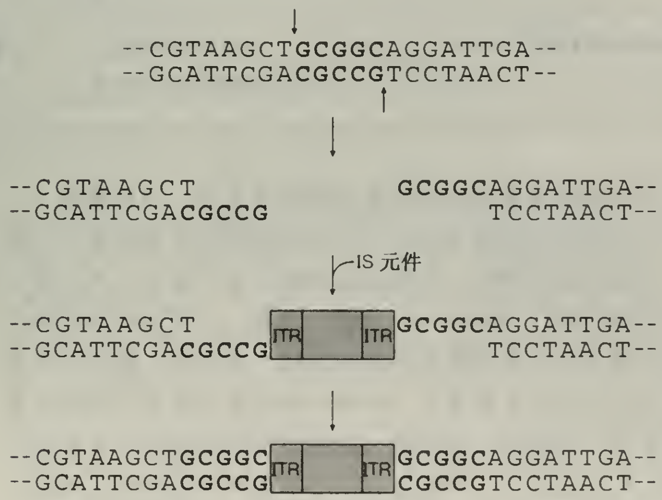


图 F4.2 插入序列转座进入宿主 DNA，并形成目标位点的副本（粗体字所示），ITR，指末端反向重复序列

许多真核生物转座子有着与反转录病毒基因组（参见 R4）相似的结构，酵母中 6kb 的 **Ty 元件**（在每个基因组中大约有 30 个全长拷贝）编码的蛋白与反转录病毒的反转录酶和整合酶相似，Ty 元件两侧为**长末端重复序列**（LTR），它先转录成 RNA 再反转录到 DNA 双螺旋中，然后插入到其他地方。*Drosophila* 中的 *copia* 元件（5kb 序列，50 拷贝）在结构上很相似，这种成分称作**反转录转座子**。一般认为反转录病毒就是那些可存在于细胞之外并转移到其他细胞的反转录转座子。在高等真核生物中发现的一些分散重复序列（如 LINES 和 SINES）很可能是通过转座作用在基因组内传播的（参见 D4）。LINES 与反转录病毒基因组相似，而作为 SINE 的 *Alu* 元件则与**信号识别颗粒**中的 **7SL RNA** 有很大的相似性，这种颗粒是可能起源于这种 RNA 反转录物，并与内质网中新合成的多肽分泌相关的一种复合物（参见 Q4）。

（王薛林 译 刘进元 校）

G1 DNA 克隆概述

要 点

DNA 克隆	通过将生物体基因组 DNA 片段作为自主复制载体的一部分进行独立复制的方式，使对该片段的分离及操作简单化。
宿主和载体	基因克隆过程中的大多数常规操作中都使用大肠杆菌作为宿主细胞。质粒和噬菌体可作为大肠杆菌的克隆载体。质粒、病毒及整个染色体的载体一直被用作外源基因导入其他原核及真核生物体内的工具。
亚克隆	亚克隆只是简单地将已被克隆了的 DNA 片段从某一载体向另一载体的转换；服务于基因克隆中的许多常规操作技术。
DNA 文库	由基因组或 cDNA 的一套随机克隆片段构成，其中每个片段连在一个单独的载体分子上，常用来分离未知基因。
筛选文库	通过与蛋白质产物推测出的 DNA 序列或某一相关基因序列的杂交，或者通过克隆片段的蛋白产物来筛选基因的技术。
克隆分析	鉴定后的已克隆基因可用限制性酶切图谱进行分析，并最终进行 DNA 序列分析。然后再将该基因序列用于 DNA 克隆的各种应用。

相关主题 克隆载体（H） 克隆 DNA 的分析与应用（J） 基因文库与筛选（I）

DNA 克隆 长期以来，由于生物体内某些蛋白质或其他组分的含量稀少难以大量纯化，造成对这些物质的详尽分子分析极其困难或者几乎不可能。一种途径是分离负责某一蛋白质表达的或某一产物形成的基因。然而每个生物体的基因组都很庞大、复杂（参见 D），且任一目的基因序列通常在单个细胞中只出现一到两次，因此标准的化学或生化方法都不能被用来分离基因组中的特定区域以对其进行研究，尤其是当目的 DNA 序列与所有其他序列具有化学相同性时更加困难。解决这一难题的方法是将基因组中携有该基因或其他相关序列的较小片段连接到一段可以自主复制的 DNA 即**载体**上，形成通常可以在另一宿主中进行复制的**重组 DNA**，这种复制是独立于原初基因组的。带有重组 DNA 的宿主细胞的增殖构成了一群具有遗传一致性的个体，或称为**单克隆**。这一系列操作过程就被称作 **DNA 克隆**。

在进行 DNA 克隆中，通常将以下操作、制作或生产归为**遗传工程**范畴：

- DNA 序列分析，以及由此派生的蛋白质序列分析（参见 J2）。
- 对基因启动子及其他调控序列的分离与分析（参见 J4）。
- 通过大量正常和突变形式的产物研究来了解这些蛋白质/酶/RNA 的功能（参见 J5）。
- 突变的鉴定，例如由于基因缺陷导致形成的疾病（参见 J6）。
- 生物技术，如蛋白质及其他重要生物功能的分子如人胰岛素和生长素的大规模工业化生产（参见 J6）。
- 工程动物、工程植物以及基因治疗（参见 J6）。
- 改变了特性的工程蛋白质（参见 J6）。

宿主和载体

对 DNA 片段最初的分离及分析几乎都是以大肠杆菌为**宿主**进行的，虽然酿酒酵母（*Saccharomyces cerevisiae*）被用来操作人类基因的大片段（参见 H3）。很多种天然复制子具有用作**克隆载体**所需的特性。为了使某些获得克隆基因的长期表达，一些载体被设计用来将外源 DNA 整合到宿主细胞基因组，但载体自身通常必须能够独立于宿主细胞基因组之外进行复制与分离。载体还应包含一个**选择标记**，即一个能使含有该载体的宿主可从那些不含该载体的细胞中被筛选出的基因，一般是赋予宿主对某种毒素的抗性（参见 G2），或使宿主细胞能在特定生长条件下生存（参见 H3）。

最早的大肠杆菌载体是染色体外（与染色体相独立的）的环状**质粒**（参见 G2），以及许多**噬菌体**（侵染细菌的病毒；参见 R2）。与质粒载体相比，**λ 噬菌**体可用来克隆较大的 DNA 片段，噬菌体 M13 可以使克隆 DNA 以单链形式被提取出来（参见 H2）。许多专用载体被改造成兼具质粒与噬菌体两者特性的形式，比如质粒-λ 噬菌体杂合体即**黏粒**（参见 H3）。人类和其他物种的巨大基因组片段已被用**细菌人工染色体**（bacterial artificial chromosome，BAC）的形式克隆进大肠杆菌，或以**酵母人工染色体**（yeast artificial chromosome，YAC；参见 H3）克隆进酿酒酵母。

质粒和噬菌体载体已被用来在大肠杆菌外的一些细菌中表达外源基因，而且一些噬菌体还可被用来将 DNA 整合到宿主细胞基因组中，如 λ 噬菌体（参见 H2）。已经开发出能在酵母中应用的质粒载体（**酵母游离型质粒**，yeast episomal plasmid），而对于植物一种细菌质粒（根癌农杆菌 **Ti 质粒**，*Agrobacterium tumefaciens* Ti plasmid）可被用来将 DNA 整合到基因组中。对于培养状态的其他真核细胞，常用那些能自然侵染目标物种的病毒作为载体将其 DNA 维持在染色体外，或者是整合到宿主基因组（例如 **SV40**、**杆状病毒**、反转录病毒；参见 H4）。

亚克隆

是克隆实验中最简单的一种，可以此来说明 DNA 克隆中的许多基本

技术，即将已克隆的 DNA 片段从一载体向另一载体的转移，这一过程被称为**亚克隆**。举例来说，亚克隆可用来对较大的克隆片段上较短区段进行仔细研究，或是将基因转移到能在特定物种中表达的另一载体上。就大肠杆菌的质粒载体来说，最常见的亚克隆程序可分为以下步骤，这在 G2～G5 中有更详尽的讨论：

- 含有目标克隆序列的质粒 DNA 的**提取**（参见 G2）。
- 用限制性内切核酸酶将质粒**酶切**成不连续的片段（参见 G3）。
- 经**琼脂糖凝胶电泳分离**片段（参见 G3）。
- **纯化目标**片段（参见 G3）。
- 将目标片段**连接**在一新质粒载体上，形成新的重组分子（参见 G4）。
- 将连接好的质粒转入某一大肠杆菌菌株（**转化**，transformation）（参见 G4）。
- 转化细菌的**筛选**（参见 G4）。
- 重组质粒的**分析**（参见 G4）。

DNA 文库

用于鉴定未知基因的克隆实验的 DNA 有两个主要来源：目的物种的基因组 DNA 和真核生物特定基因表达细胞或组织的 mRNA 群体，分别被用于构建**基因组文库**和 **cDNA 文库**（参见 I1 和 I2）。

DNA 文库是一套 DNA 克隆（一个克隆是一个遗传上独特的个体或一群相同的个体），每个克隆都是插入了不同片段的载体在宿主中扩增后的产物。基因组文库是用基因组 DNA 的随机片段制备成的。然而用基因组文库来克隆某一基因是一种效率极低的方法，特别是对于庞大的真核生物基因组，其中大量 DNA 是非编码的（参见 D4）。另一种是用来自表达目的基因细胞或组织的 mRNA 作为来源构建成的文库即 cDNA 文库。经**反转录**将 mRNA 合成 cDNA（DNA 拷贝）后，插入载体构建成 cDNA 文库。用 cDNA 文库来克隆目的基因是有效的，但克隆的仅是其编码区，而不包含编码区以外的基因组序列。

筛选文库

因为文库中的哪一个克隆含有目的基因并不是从外表可以看出的，所以必须用一种方法筛选出含有目的基因的克隆。通常是用一段与目的基因序列的某一区段互补或部分互补的放射性或荧光标记的 DNA **探针**，通过**杂交**来检测出该基因（参见 C3）。若基因的蛋白产物可获得，则探针可以是从该蛋白产物序列推测出的一段寡核苷酸，或者探针也可以来自另一物种中的某一相关基因（参见 I3）。制备探针越来越常用的一种方法是**聚合酶链反应**（PCR；参见 J3）。另一筛选方法是基于文库中被克隆的编码区表达后，对其蛋白产物进行活性鉴定，或是用特异抗体对其进行鉴定（参见 I3）。

克隆分析

一旦含有目的基因的克隆被确定，就可以通过**限制性酶切图谱**即用限

制性酶切来分析 DNA 片段，来对其克隆的片段的结构进行更进一步的研究（参见 J1），以至最终对全长片段进行测序（参见 J2）。随后可通过将所得全长序列与数据库中其他已知序列进行比较分析，确定出该蛋白产物的完整序列（参见 J2）。至此就可以用如上所述的任何克隆程序对该序列进行操作。

在体外进行 DNA 克隆与分析要用到许多酶，其中常用酶的特性特列于表 G1.1 中，详细描叙其应用的章节也已列于其后。

表 G1.1　用于 DNA 克隆的酶

酶	用　途
碱性磷酸酶	从双链或单链 DNA，或 RNA 的 5′端移去磷酸基（参见 G4）。
DNA 连接酶（来自噬菌体 T4）	由 ATP 水解提供能量，将 dsDNA 的戊糖-磷酸骨架的 5′-磷酸与 3′-羟基相结合。待连接的 DNA 末端必须是相匹配的，即平端与平端，或为互补的黏性末端（参见 E2 和 G4）。
DNA 聚合酶 Ⅰ	由含游离 3′-羟基的引物起始，沿 5′至 3′的方向合成与 DNA 模板互补的 DNA 链（参见 E2）。Klenow 片段是 DNA 聚合酶 Ⅰ 的缺失 5′至 3′外切核酸酶活性的截短部分。
外切核酸酶 Ⅲ	外切核酸酶从线性 DNA 的末端依次切除核苷酸。外切核酸酶 Ⅲ 只从 dsDNA 的 3′端切割（参见 J4）。
绿豆核酸酶	降解单链核酸，留下完整的双螺旋区段（参见 J4）。
核酸酶 S1	类似绿豆核酸酶，并且能够降解对应于互补链缺刻对面的单链（参见 J4）。
多核苷酸激酶	一个依赖 ATP 的反应，向双链或单链 DNA/RNA 的 5′-羟基末端添加磷酸基。若使用［γ-^{32}P］ATP，则可将 DNA 标记上放射性同位素（参见 I3）。
限制酶	在识别序列处（通常对称）切割 dsDNA 的双链。水解戊糖-磷酸骨架，使其一端为 5′-磷酸基而另一端为 3′-羟基，形成平端或“黏”末端（5′-或 3′-凸端）（参见 G3）
反转录酶	依赖 RNA 的 DNA 聚合酶。由含有游离 3′羟基的引物起始，沿 5′至 3′方向合成与 RNA 模板互补的 DNA 链，需要 dNTP（参见 I2）。
RNase A	只降解 RNA，而不降解 DNA 的核酸酶（参见 G2）。
RNase H	降解 RNA-DNA 异源双链中的 RNA 链的核酸酶（参见 I2）。
T7，T3 和 SP6 RNA 聚合酶	分别由相应的噬菌体编码的专一性 RNA 聚合酶。每种酶只能识别自身的噬菌体 DNA 的启动子，特异地转录各自启动子下游 DNA 序列（参见 H1 和 I3）。
***Taq* DNA 聚合酶**	来源于嗜热细菌（*Thermus aquaticus*）的 DNA 聚合酶。最适温度为 72℃，在 90℃ 以上仍相当稳定。用于 PCR（参见 J3）。
末端转移酶	可将核苷酸加到线性单链或双链 DNA 或 RNA 的 3′端。如只用 GTP，则仅加上多个 G（参见 I2）。

G2 质粒 DNA 的制备

要 点

载体质粒	细菌质粒是独立于宿主基因组复制、编码抗生素抗性基因的小型环状 DNA 分子、运载克隆 DNA 的常用载体。
质粒的小量制备	仅从几毫升培养液中提取小量质粒用作分析，这一过程被称作小量制备（minipreparation 或 miniprep)。
碱裂解	用 SDS 碱性溶液裂解大肠杆菌细胞，变性蛋白质和 DNA，然后中和沉淀染色体 DNA 以及绝大部分蛋白质，在溶液中只剩下质粒 DNA 和 RNA。
酚抽提	用酚或酚-氯仿混合液抽提以去除碱裂解液中残余蛋白质。
乙醇沉淀	加入乙酸钠和乙醇后离心，可从溶液中沉淀出核酸。这一方法常用来浓缩样品。
氯化铯梯度	CsCl 梯度离心是质粒大量制备过程中的一步，可将超螺旋质粒 DNA 从蛋白质、RNA 及线性或带缺刻的 DNA 中纯化出来。
相关主题	DNA 克隆概述（G1） 连接、转化与重组体分析（G4） 基因文库与筛选（I）

质粒载体

20 世纪 70 年代中期，最早使用的克隆载体是来自大肠杆菌的未加工的细菌**质粒**。质粒为**染色体外**的小型环状分子，大小约为 2～200kb，通常以多拷贝（可多达几百个）形式存在于宿主细胞大肠杆菌中。虽然质粒的正常复制依赖宿主细胞中的聚合酶及其他成分，但质粒含有**复制起点**（*ori*；参见 E1）用以保证质粒的自主复制。质粒一般仅携有几个基因，其中之一会赋予细菌对抗生素物质的**抗性**。最常见的抗性基因是 *bla*，或 amp^r 基因，编码可以降解青霉素类抗生素如**氨苄青霉素**的 β-内酰胺酶。另一种常见抗性基因是 *tetA* 基因，编码一种可以从细胞内将四环素类抗生素排出的跨膜泵蛋白。

质粒的小量制备

如将一个基因从某一质粒向另一质粒转移（参见 G1）的亚克隆过程的第一步就是要提取质粒 DNA。由于质粒远远小于大肠杆菌染色体 DNA（参见 D1)，可以用物理化学方法将它们相分离，例如碱裂解法（参见下

文)。从几毫升菌液中提取出足够量的质粒 DNA 供初步操作所需是完全可以的，这样的提取操作通常称为**小量制备**。将携有目的质粒的大肠杆菌菌株接种于数毫升液体培养基，待增殖至稳定期（过夜培养）后，离心菌液形成菌体沉淀，接着就可以用小量制备法提取质粒。

碱裂解

从染色体 DNA 及大部分其他细胞成分中纯化出质粒 DNA 最常用的方法称为**碱裂解法**（图 G2.1)。菌体沉淀重悬于悬浮缓冲液，必要时该缓冲液可加入能降解菌体细胞壁的溶菌酶。再加入细胞**裂解液**，即含去污剂十二烷基硫酸钠（SDS）的碱性 NaOH 溶液。SDS 的作用是破碎细胞膜（参见 A1）及引起蛋白质变性；碱性条件会引起 DNA 变性及水解 RNA（参见 C2)。然后用高浓度 pH5 的乙酸钾溶液**中和**。其作用是沉淀已变性蛋白质和染色体 DNA 及大部分去污剂（十二烷基硫酸钾不溶于水)。将溶液再次离心，上清（**裂解液**）中含有质粒 DNA。小型闭合环状的质粒 DNA 在碱处理后很易复性，此时的溶液中还含有大量小分子 RNA 及少量蛋白质。

酚抽提

目前有许多专利方法用以从上述裂解液中提取出纯质粒 DNA，其中多数是采用将 DNA 选择性地结合于树脂或膜，再洗脱掉蛋白质和 RNA 的原理。而经典的方法，虽然较慢却极为有效，采用对裂解液用**酚**或**酚-氯仿**混合液进行抽提。酚相与水相不相溶，当剧烈振荡然后静置或离心分相后，变性了的残余蛋白会沉降出来处于酚相与水相的界面（图 G2.1)。

乙醇沉淀

留在水相中的 DNA 和 RNA 可由**乙醇沉淀**得以浓缩。这是常用程序，可应用于任何核酸溶液。当向溶液中加入 NaAc 至 Na^+ 浓度大于 0.3 mol/L，溶液中的 DNA 和（或）RNA 加入 2～3 倍体积的乙醇后而被沉淀出来。离心沉淀核酸然后再溶于更少的体积，或溶于含新成分的缓冲液中。对于小量制备过程，乙醇可直接加到酚抽提后的裂解液中，形成的沉淀可溶于 Tris-EDTA 缓冲液，该缓冲液是用于 DNA 贮存的常用溶液。溶液中含 Tris·HCl 以缓冲溶液酸碱度（通常为 pH8)，还含有低浓度的 EDTA 用来螯合 Mg^{2+} 以保护 DNA 免受核酸酶降解，因大多数核酸酶工作都需要 Mg^{2+}。**核糖核酸酶 A**（RNase A）可同时被加入溶液中以降解残余的 RNA 污染。该酶降解 RNA 又不损伤 DNA，也不需 Mg^{2+} 参与反应。

氯化铯梯度

碱裂解法也可应用于大量制备获得毫克数量级的质粒 DNA。制备大量的质粒可作为常用克隆载体贮备，或用作大量酶解反应的底物。此时 **CsCl 密度梯度离心**（参见 C2）可被用来作为最终纯化步骤。该步骤虽然有点费力，但却是获得极纯的超螺旋质粒 DNA 的最佳方法。粗裂解液经含有溴化乙锭（参见 C4）的 CsCl 梯度（参见 C2）分级分离后，各种成分会完全分离（图 G2.1)。超螺旋 DNA 结合的溴化乙锭比线性或带切口的

要小（参见 C4），具有较大的密度（因其损伤最少）。因此大量的超螺旋质粒可一步就从蛋白质、RNA 及污染的染色体 DNA 中分离出来。从离心管中回收含有质粒带的溶液，在除去溴化乙锭后再用上述乙醇沉淀步骤进行浓缩。

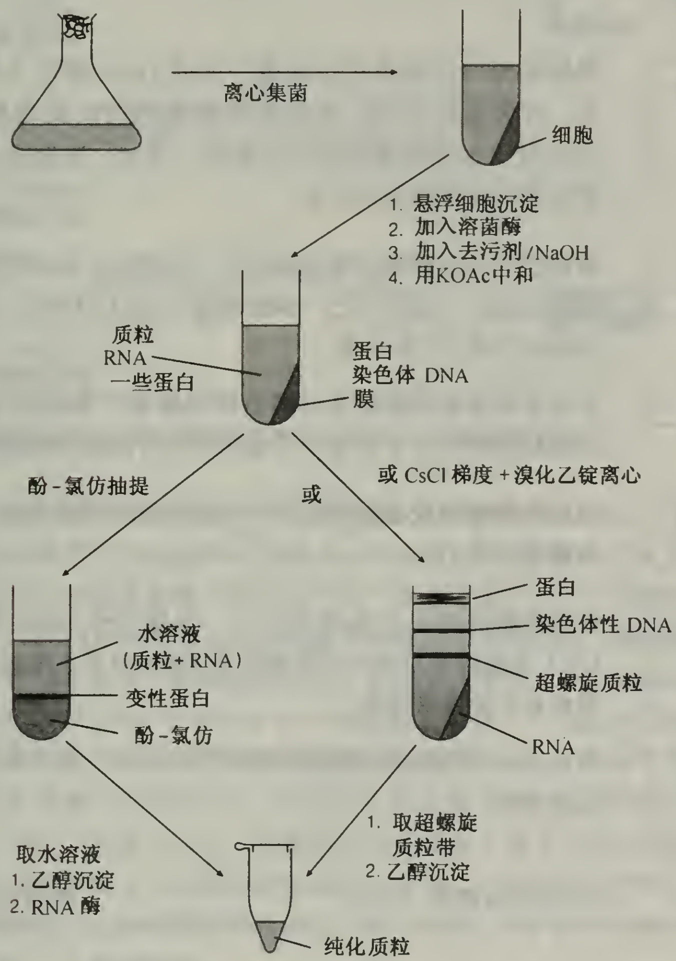


图 G2.1　质粒的制备流程（详见正文）

G3 限制酶与电泳

要 点

限制性内切核酸酶	限制性内切核酸酶来自细菌，可将 DNA 酶切（水解）成特定的、可再生的片段。在细菌中限制性内切核酸酶是抵抗外源 DNA 侵入的限制修饰防卫机制的一部分。限制性内切核酸酶是进行基因克隆的基本工具。
识别序列	限制性内切核酸酶在很短的回文（对称的）识别序列处对称切割 DNA 双链，产生一个 5′-磷酸基和一个 3′-羟基，形成平末端或突出的 5′或 3′-黏末端。
黏末端	含有单链末端的限制性内切核酸酶酶解产物被称为“黏”末端，因此黏末端可与带有互补末端的其他片段进行碱基的退火配对。
限制性酶解	在用琼脂糖凝胶电泳对目的片段进行分析或纯化前，商品酶被用来酶解质粒 DNA。
琼脂糖凝胶电泳	琼脂糖凝胶可以在电场作用下，根据线性 DNA 分子的大小及 DNA 在凝胶基质中的迁移率分离 DNA 分子。电泳还可用于测定质粒分子的大致构型。
DNA 片段的分离	特定 DNA 片段可从琼脂糖凝胶中切出后，回收纯化用于随后的克隆实验。

相关主题 DNA 克隆概述（G1） 基因文库与筛选（I）

限制性内切核酸酶

为了将外源 DNA 片段插入质粒载体，需要对双链 DNA 进行酶切和再连接。20 世纪 60 年代和 70 年代初，对**限制性内切核酸酶**的鉴定及操作是实现克隆 DNA 的关键。许多细菌中含有限制修饰系统，构成了对侵入细胞的外源 DNA 的防卫机制。该系统由两个主要组分构成，一个是限制性内切核酸酶，可识别一小段对称的 DNA 序列（图 G3.1），然后在识别位点处切割（**水解**）双链 DNA 骨架，外源 DNA 因此被降解成相当短的片段。该系统的第二个组分是**甲基化酶**，该酶可以对细胞 DNA 的相应识别序列内的 C 或 A 碱基上加一个甲基（参见 C1）。这种修饰作用可以保护宿主细胞 DNA 不被内切酶降解。

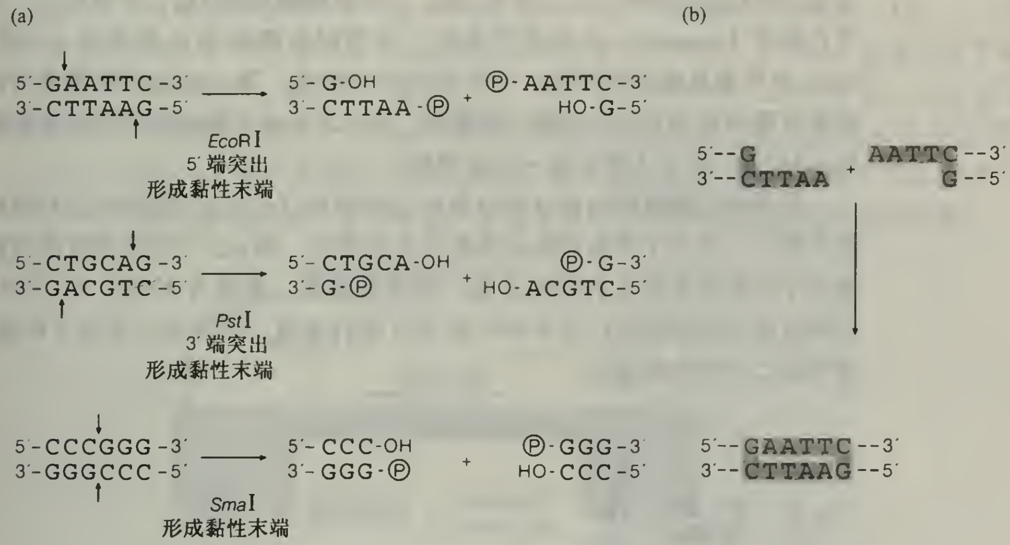


图 G3.1 (a) 限制性内切核酸酶在其识别序列处的作用；(b) 黏末端的退火

识别序列

限制性内切核酸酶（简称**限制酶**）的作用方式描绘于图 G3.1a 中，其中包括作为示例的原始型酶 *Eco*RⅠ。这种以二聚体形式起作用的酶仅识别 6bp 的**回文序列**（即双链中每一条链均按 5′→3′方向阅读，其序列相同），线性 DNA 上该位点处的酶切形成两个双链片段（**限制性片段**），每一片段都具有同一的磷酸基团的 5′单链突出末端，而 3′端则是游离羟基。在随机序列 DNA 中平均每 $4^6=4096$bp 长存在一个 6bp 的识别序列，因此一个巨大的 DNA 分子可被这种酶切割成平均长 4kb 的特异片段。目前已知的限制酶已达数百种，且大部分可以购买到。这些酶的识别序列在 4bp～8bp，或更长一些，其酶切产物是具有 5′或 3′突出的末端或平末端。新形成的 5′端总是保留着磷酸基团。图 G3.1 还包括另外两个例子。限制酶对酶切位点的极高特异性保证了将大片段 DNA 分子及载体切割成的特定片段，且重复性好。

黏末端

具有突出末端的限制性酶解反应的产物的末端被称为“**黏合的**（cohesive）”，或“**黏**（stick）”末端，其特性在于它们可通过单链尾端的碱性配对（退火），与其他任何具有相同突出序列的末端结合。例如任何一个由 *Eco*RⅠ酶切产生的片段均可与相同方式产生的任何其他片段退火（图 G3.1b），然后通过连接共价结合在一起（参见 E4）。事实上，在某些情况下，由不同识别序列的酶所形成的 DNA 末端也可以是相同的，产生能够相互配对的单链末端。

限制性酶解

按不同的分析或制备目的，用商品化的限制酶及其缓冲液对质粒或基

因组 DNA 进行酶解反应（参见 I1）。所有限制酶均需 Mg^{2+} 参与反应，通常在高至 10mmol/L 的浓度下进行，但不同的酶有各自最适的 pH 值、NaCl 浓度或其他溶液成分以达到最佳反应活性。某一特定的酶所需的缓冲液以浓缩物的形式与该酶一起提供。图 G3.2 表示用两种不同的限制酶 *Bam*HⅠ和 *Eco*RⅠ酶切同一样品质粒。

若用琼脂糖凝胶电泳分析只要对几百纳克（<1μg）的质粒进行酶解就足够了；若为了制备可能会需要几毫克质粒。如 G2 中所述前者的用量相当于小量制备样品的百分之几。在最适温度（通常为 37℃）下，DNA 与酶及相应的缓冲液以总体积约为 20μl 进行保温，然后加入染料上样液，将样品上样琼脂糖凝胶。

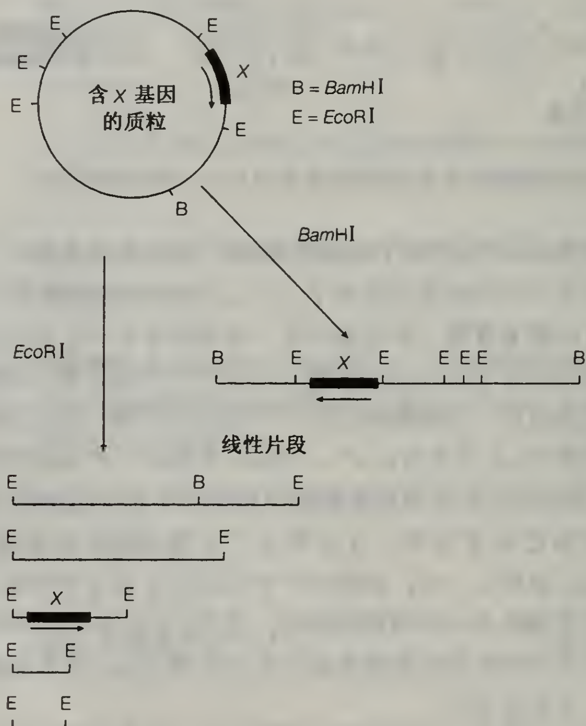


图 G3.2 两种不同的限制酶酶解同一质粒

琼脂糖凝胶电泳

琼脂糖是从海藻中提取的一种多糖，当以 0.5%～2% 的浓度（w/v）溶解于水溶液后可形成固体凝胶。用于电泳的琼脂糖是比用于制备细菌培养平板琼脂的更纯产品。

当电场加载于具导电性的缓冲液中的琼脂糖凝胶上后，DNA 片段在凝胶中依据其大小和形状以一定的速率向正极方向移动（DNA 具有很高的负电性，参见 C1）（图 G3.3）。较小的线性片段比较大片段移动得快，因为较大片段会被构成凝胶的琼脂糖纤维网的障碍所阻滞。因此**电泳**过程

可被用来根据大小分离不同的 DNA 片段。凝胶的不同浓度〔1%、1.5%(w/v) 等〕在不同大小范围内有各自最佳分辨率。将 DNA 样品上样于凝胶表面的**上样孔**内（图 G3.3），打开电源开关后，DNA 将沿不同的泳道或路径在凝胶中迁移。加入的染料也同样迁移，被用来指示电泳进程。DNA 可通过在凝胶中加入**溴化乙锭**（参见 C4），或将凝胶在电泳后浸于溴化乙锭溶液中进行染色，在紫外光照射下 DNA 呈现出橙色的条带。

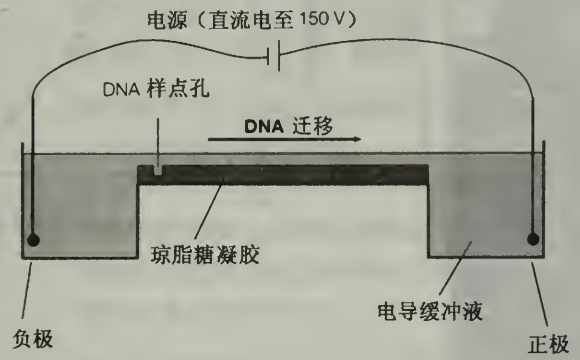


图 G3.3　琼脂糖凝胶装置

图 G3.4a 描述了图 G3.2 所示酶解反应形成片段的凝胶电泳结果。未酶解的质粒（泳道 U），与 *Bam*HⅠ（泳道 B）及 *Eco*RⅠ（泳道 E）酶解后的样品同时电泳。一套已知大小的线性标准 DNA 片段（泳道 M）以两种不同浓度上样于样品的两边，其大小已标于图中。应注意以下几点：

- 琼脂糖凝胶电泳中未酶解的质粒 DNA（泳道 U）通常呈现两条带。较低位置、迁移快的条带是从细胞中提取的完整负超螺旋质粒 DNA。因其构型紧密故具有较高迁移率（参见 C4）。较高位置的条带是由于超螺旋 DNA 的一条链被断裂而形成的开环或有切口 DNA 构成，因具开环构型，故迁移率较低。
- DNA 酶解后清楚地显示出单一片段（泳道 B）和五个片段（泳道 E），片段大小可通过与标准 DNA（泳道 M）相比较来估算。泳道 E 中条带的亮度与片段的大小成比例，因为摩尔浓度一定时，小片段含有较小的 DNA 分子质量。这对于分子质量标准物也适用，这里的标准物是由 48.5kb 的 λ 噬菌体线性 DNA 分子酶解产生的（参见 H2）。泳道 U、B、E 中所含 DNA 的量是不相等的，这里的 DNA 用量是能将所有片段清晰显示出来的最适用量。
- 要精确测定线性片段大小，可通过泳道 M 中已知片段大小的对数对每条片段的迁移距离作标准曲线，再从曲线上取点计算来进行。图 G3.4b 是一条相当直的线，通常对于大片段会产生偏差。根据片段在同一凝胶中的迁移率，从图中读出其分子大小的对数，可以得出未知线性片段的大小。用这一方法不能测定未酶解环状质粒的大小，因为

环状与线性 DNA 在凝胶中的相对迁移率取决于电泳条件（温度、电场等）。

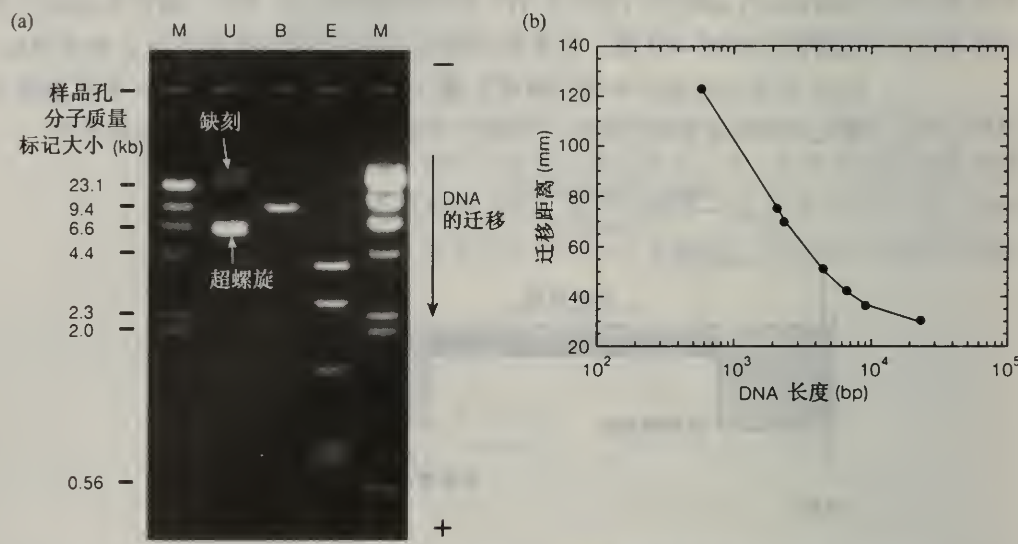


图 G3.4　(a) DNA 限制酶解片段的琼脂糖凝胶电泳（详见正文）；(b) 迁移距离对片段大小所作的标准曲线

DNA 片段的分离

琼脂糖凝胶也可用于分离制备特定 DNA 片段，为随后的连接反应或其他克隆实验提供样品。将片段从凝胶中切出，选用一种方法，从溴化乙锭染色的琼脂糖中分离纯化出 DNA。假设含基因 *X*（图 G3.2）的 *Eco*R Ⅰ片段是亚克隆的目标 DNA（参见 G1），则应将图 G3.4a 泳道 E 的最大片段从凝胶中纯化出来，用来与新载体相连接（参见 G4）。

G4 连接、转化与重组体分析

要 点

DNA 连接	T4 DNA 连接酶可以修复双链 DNA 骨架的断裂，催化限制酶酶解反应的逆反应，即使退火的互补黏性末端共价连接在一起，形成新的 DNA 分子。
重组 DNA 分子	限制酶及随后 DNA 连接酶的使用，可以产生目的 DNA 片段插入载体质粒的重组质粒。
碱性磷酸酶	用碱性磷酸酶处理线性载体分子，以除去 5′-磷酸基团，使载体在没有目的片段插入的情况下不能连成环状，从而降低反应物中载体自身环化的比例。
转　　化	转化是细菌吸收外源 DNA，通常是质粒 DNA 的过程。质粒通过转入具有特定遗传性状的大肠杆菌菌株而被克隆。用 Ca^{2+} 处理大肠杆菌细胞，可使之成为易接受质粒 DNA 的感受态细胞。将这样的细胞涂布于琼脂平板，会长出单克隆或多个克隆。
筛　　选	吸收了质粒的细菌可以通过在含某种抗生素的平板上生长的方法进行筛选，该质粒载体上携有编码对所用抗生素抗性的基因。
转化率	转化率定义为每毫克转化的质粒 DNA 所产生的具抗生素抗性的克隆数。
筛选转化子	多数例子中，如用 DNA 文库、质粒和其他载体被设计成利于筛选重组质粒的转化子，而对于简单的亚克隆实验，常将小量制备的转化子 DNA 进行酶解后，经琼脂糖电泳分析来筛选转化子。
转化子的增殖与保存	从转化平板上挑取单克隆，接种于含有选择标记抗生素的液体培养基中培养，然后将部分菌液冻存于含一定浓度甘油的保存液中备用。
凝胶分析	重组质粒与原载体可经琼脂糖凝胶电泳，因大小不同而相区分，也可用插入时所用的限制酶再将目的片段切割出来。
插入片段定向	插入片段在载体中的方向可使用某种已知能不对称切割插入片段的限制酶酶解该质粒，再经琼脂糖凝胶电泳分析来确定其插入方向。
相关主题	DNA 克隆概述（G1）　　　　克隆载体（H）

DNA 连接

为了将目的 DNA 片段插入载体，必须有一种能共价连接 DNA 分子的方法，DNA 连接酶正好能执行此功能。**DNA 连接酶**可以修复（**连接**）双链 DNA 分子中一条链上的断裂（参见 E2），所提供的 5′-末端须携有磷酸基团。连接反应需要腺苷酸化的试剂来活化磷酸基团以利于 3′-羟基作用，大肠杆菌的连接酶用 NAD^+，而最常用的来自 T4 噬菌体的酶是以 ATP 为腺苷酸化的试剂。连接酶可有效地在黏性末端退火碱基对处闭合断裂的磷酸二酯键（参见 G2）。连接反应本质上是限制酶酶解反应的逆反应，T4 连接酶甚至可以连接平末端，但效率较低。

重组 DNA 分子

我们设计一个实验，即将一个含目的基因（*X*）的 DNA 片段插入进质粒载体（图 G4.1）。目的 DNA 可以是从琼脂糖凝胶中回收的单一片段（参见 G3），或是来自如基因组 DNA 的多个片段的混合物（参见 I1）。如果目的片段是由 *Eco*R Ⅰ 酶解所制备的，那么该片段可以与用相同酶降解后的载体 DNA 连接（图 G4.1）。实际上载体上应该只含有相关酶的惟一切割位点，否则正确的产物须由三个甚至更多个片段连接而成，那将是效率非常低的。这种连接反应会产生多种可能的产物，最终结果取决于不同片段的相对浓度及反应条件，但目的产物将是目的片段插入在载体分子

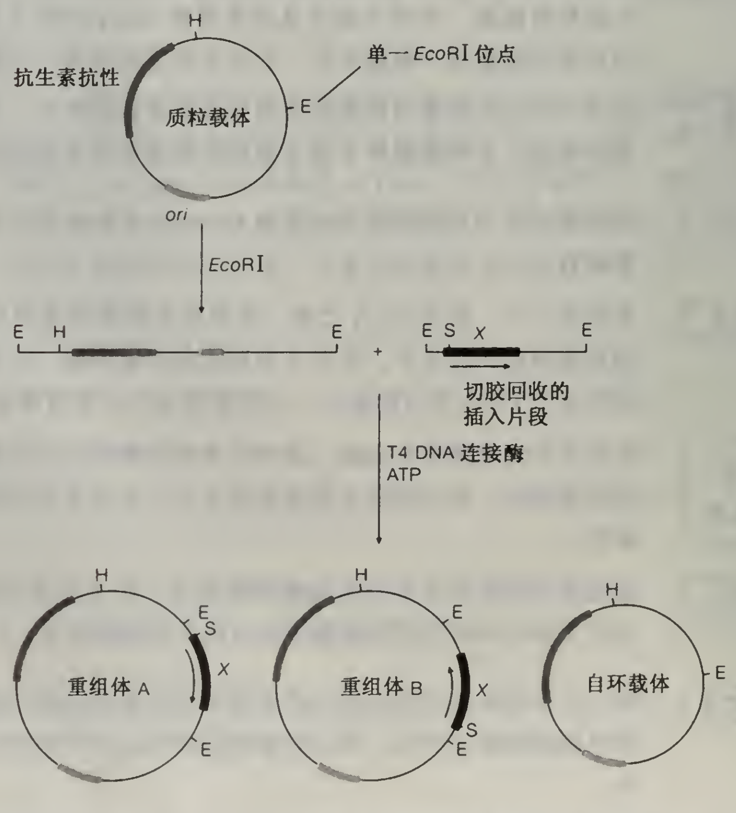


图 G4.1 载体与目的片段连接形成重组体和非重组产物的示意图

*Eco*RⅠ酶切位点处（正反插入均有）形成的环状分子，即**重组**分子（图 G4.1）。原质粒载体的再生，经该线性载体的自环化，是一种竞争性副反应，会给重组产物的鉴定带来困难。解决办法是用一对特异的限制酶酶解目的片段与载体，这样它们的两端互为不互补的黏末端，载体自身环化的可能性就会大大降低。

碱性磷酸酶

如果不方便用两种限制酶酶解，则可以在限制酶酶解反应后用**碱性磷酸酶**处理线性载体片段，该酶可从 DNA 分子的 5′端去除磷酸基。因为没有连接反应所需的磷酸基，线性载体将无法再连成环状（图 G4.2）。但与目的 DNA 插入片段的连接仍可进行，因为每个酶切位点处都有一个磷酸基可连接一条链（图 G4.2）。剩下的另一条链的切口将在转化后由细胞修复机制修补（参见下文）。

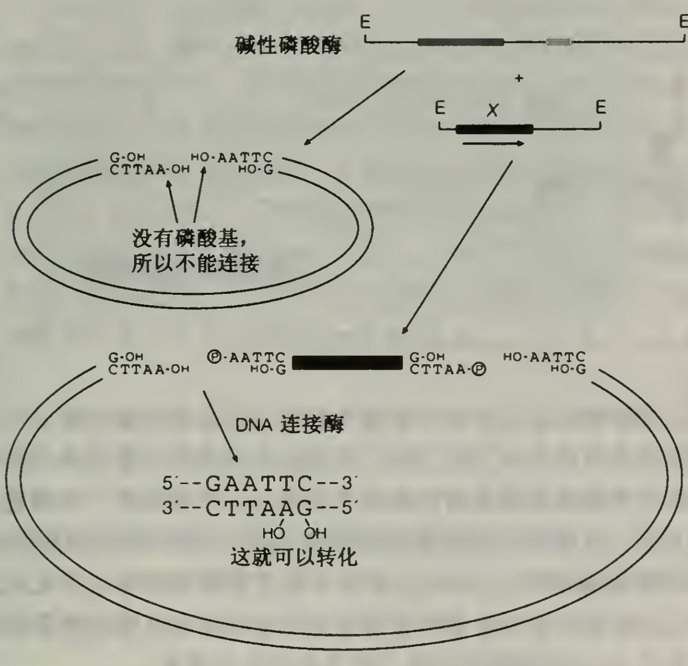


图 G4.2 使用碱性磷酸酶防止载体分子的自身环化

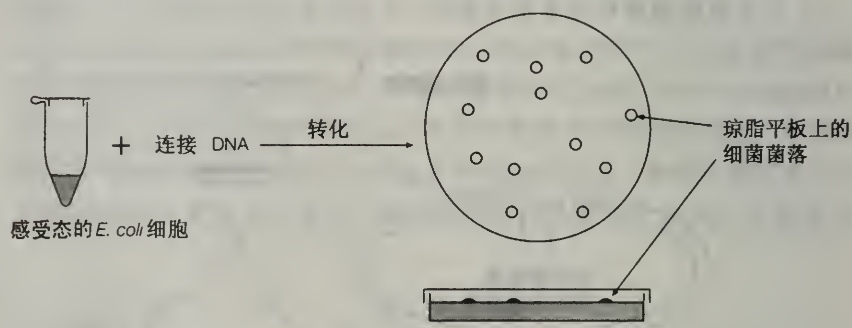
转化

连接反应中形成的重组体及其他质粒分子的混合群体必须通过转入**宿主细胞**将它们相互分离，进而进行自主复制（**克隆**）。目前单一的克隆实验中最常用的宿主细胞乃是具有特定遗传特性的大肠杆菌菌株。所谓特定遗传特性，例如最基本的条件就是这些菌株不能表达限制修饰系统（参见 G3）。

人们发现用含 Ca^{2+} 的溶液处理的大肠杆菌细胞会易于吸收外源 DNA，这一过程被称为**转化**。为了使其能吸收 DNA，用 Ca^{2+}（有时使用更为外来的金属离子如 Rb^{+} 和 Mn^{2+}）预处理过的细胞，被称为**感受态细**

胞（competent cell）。

大肠杆菌的转化过程是将质粒分子溶液或连接反应生成的分子混合物与感受态细胞的悬浮液混合放置一定的时间，以使得DNA能被细胞吸收。DNA进入细菌的确切机制仍不清楚。将混合溶液于42℃**热激**1～2min，会诱导涉及DNA修复的酶以及其他细胞成分的生成，有利于细胞从转化过程中的不正常状态下恢复过来，结果提高了转化效率。然后加适量生长培养基液于转化细胞，并在适温下培养，最终涂布于琼脂平板表面，培养至形成细菌单菌落（图G4.3）。同一克隆中的所有细胞均起源于某一单独个体的分裂增殖，因此所有细胞具有相同的基因型，这里不包括自发突变（参见F1），包括转化步骤中引入的质粒（换言之，即是**克隆**或**克隆群**）。



图G4.3　转化 *E. coli* 细胞后单克隆的形成

筛选

如果转化反应中的所有感受态细胞都能在琼脂平板上生长，则将会形成成千上万甚至百万的克隆。而且大多数克隆中都不含有质粒，也无法判断哪个克隆含有质粒转化将是无效率的，可见需要一种**筛选**含有质粒克隆的方法。这通常是利用质粒载体携有某一抗生素抗性基因来实现，例如β-内酰胺酶基因（amp^r），赋予对氨苄青霉素抗性（参见G2）。这样若将转化的细胞涂布于含氨苄青霉素的平板，则只有那些含有被转化质粒，从而表达β-内酰胺酶的细胞才能生存并生长增殖。

因此可以确定转化后在氨苄青霉素平板上形成的克隆都是从携有完整β-内酰胺酶基因的质粒的单个细胞增殖而成。如果转化中使用的是连接反应混合物，那么此时我们并不能确定哪个克隆中含有带有目的片段的重组质粒（图G4.1）。

转化率

感受态细胞的质量可以通过测定某一特定样品的转化率来衡量，**转化率**定义为每毫克用于转化的DNA在选择平板上所形成的菌落数。这里DNA为纯质粒，通常是在克隆实验中的载体。

转化率的变动幅度很大，可以从粗放转化程序的10^5菌落/μg到精细转化程序高于10^8菌落/μg。粗放转化程序只适合将完整质粒转入新的宿

主菌，而精细转化程序所精心制备的感受态细胞可用于文库的构建（参见I2）。约 $10^5/\mu g$ 的转化率足以满足这里所构画的简单克隆实验。

筛选转化子

一旦获得一群转化子克隆，首先要确定哪些克隆中含有插入目的片段的重组质粒。商用质粒总是被设计成有利于这一过程的进行，关于这一点将在 H1 中有详细描述。多数情况下，如 DNA 文库的筛选（参见 I3），有必要从成千上万甚至上百万的克隆中筛选出目标克隆。这可能是所有步骤中最费时的，将在 I3 中详述。在单一亚克隆实验中，要求实验设计应充分体现重组体克隆的产率，例如可采用碱性磷酸酶处理载体。在这种情况下，通常的筛选方法是分别从若干克隆中制备质粒 DNA，然后用琼脂糖电泳进行分析。

转化子的增殖与保存

从转化平板上挑取单克隆，接种于液体培养基中培养过夜，待细菌增殖至稳定期。培养液中必须含有原平板筛选转化子所用的抗生素，以维持对质粒有无的选择。部分质粒在长期无筛选压力的生长条件下会从宿主菌体内丢失，因为那些偶然丢失了质粒后能以耗能小的方式复制的细菌将会淘汰那些携带质粒的细菌。扩增的质粒可以用小量法从培养基中制备（参见 G2）。通常在这一时刻要制备培养物的保存液，方法是将培养物分散于含有一定浓度的甘油溶液中冻存，甘油的作用是保护细胞免受冰晶形成的伤害（**甘油贮液**）。这样的贮液能使同一菌株（质粒）在必要时被接种增殖并再次用于这样的贮存。

凝胶分析

重组质粒通常可根据分子大小简单地与自身环化的载体质粒相区分，进一步的鉴定是用限制酶酶切图谱。图 G4.4 显示的是图 G4.1 所示质粒分析的理想凝胶结果。对应于载体质粒和重组体的泳道已有标注，重组质粒较大的分子质量可通过与未酶解的质粒样品（泳道 U）比较而显示出来，可见超螺旋构型及切口构型的条带（参见 G3），由 *Eco*RⅠ酶解重组体得到的插入片段条带可在泳道 E 中看到。

插入片段定向

若连接反应中所用的载体及目的片段均由同种酶进行酶解（参见 G3），则插入片段可以两个方向与载体相连。例如目的插入片段含有将有被接到载体启动子的下游的基因编码区时，插入方向就十分重要。片段定向可以用一种已知不对称切割插入片段的酶及另一种能特异性切割载体的酶共同进行酶解反应来进行。如图 G4.1 和图 G4.4 所示，用切割插入片段的 *Sal*Ⅰ（S）和只切载体的 *Hin*dⅢ（H）进行**双酶切**。如图 G4.4 所示，可以看到从两方向插入（重组体 A 和 B）的不同图谱（泳道 H/S）。

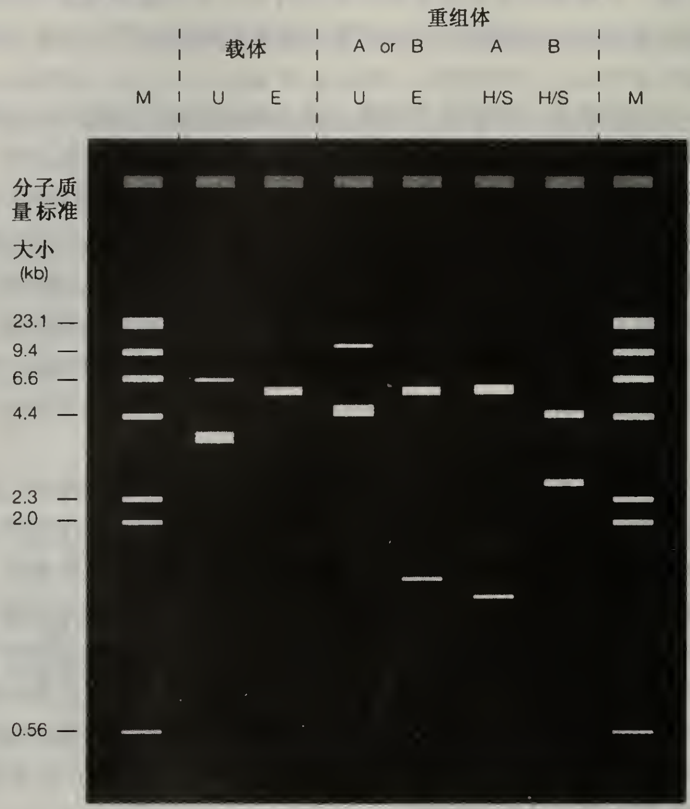


图 G4.4 琼脂糖凝胶电泳分析重组质粒（文中有详述）

（叶 彬 译 刘进元 校）

H1 质粒载体的设计

要 点

连接产物	克隆过程中最重要的步骤之一就是鉴别环化载体分子与重组质粒，已形成了许多方便于鉴别的方法。
双抗生素抗性	当载体分子上具有双抗生素抗性基因时，就可以用插入片段插入某一抗性基因而使其失活的方式来筛选重组体。
蓝-白颜色筛选	质粒上的 *lacZ'* 基因的插入失活被用来在含有 IPTG 和 X-gal 的平板上筛选重组体。当 *lacZ'* 基因完整时，IPTG 诱导其表达后，X-gal 可转化为蓝色产物，因而重组体在平板上会长成白色菌落。
多克隆位点	多克隆位点即具有多个限制酶酶切位点的一段 DNA 序列，为选择限制酶或克隆中所用的酶提供更多的选择性。
已克隆插入片段的转录	载体内的启动子可被用来在体内或体外转录插入片段。某些载体还带有两个特异启动子，从而能使插入片段的两条链均能被转录。
表达载体	许多载体被设计成使插入片段内的基因可从强启动子起始转录并表达。在某些情况下，如用 T7 表达载体，目的产物可占大肠杆菌细胞总蛋白的大部分。
相关主题	基因操作（G）　克隆 DNA 的分析与应用（J） 基因文库与筛选（I）

连接产物

将目的片段与质粒载体连接时，最常见的非目标产物就是线性载体片段自身环化所形成的空质粒载体（参见 G4）。自环化载体可通过从转化克隆中小量提取质粒，然后酶解及随后的琼脂糖凝胶电泳分析来与重组体相区分（参见 G4），但用此法进行大规模筛选是极不方便的，目前已开发出更有效的、便于筛选的一系列载体。

双抗生素抗性

pBR322 是最早开发出的质粒载体之一。pBR322 及其衍生物包含两个抗生素抗性基因，amp^r 和 *tetA*（参见 G2）。如果目的 DNA 片段插入到这两个基因中的任一编码区内，比如 *tetA*，则该基因会发生**插入失活**，而后就可以通过由转化子所显示的抗生素抗性来鉴定出重组体（图 H1.1a）。

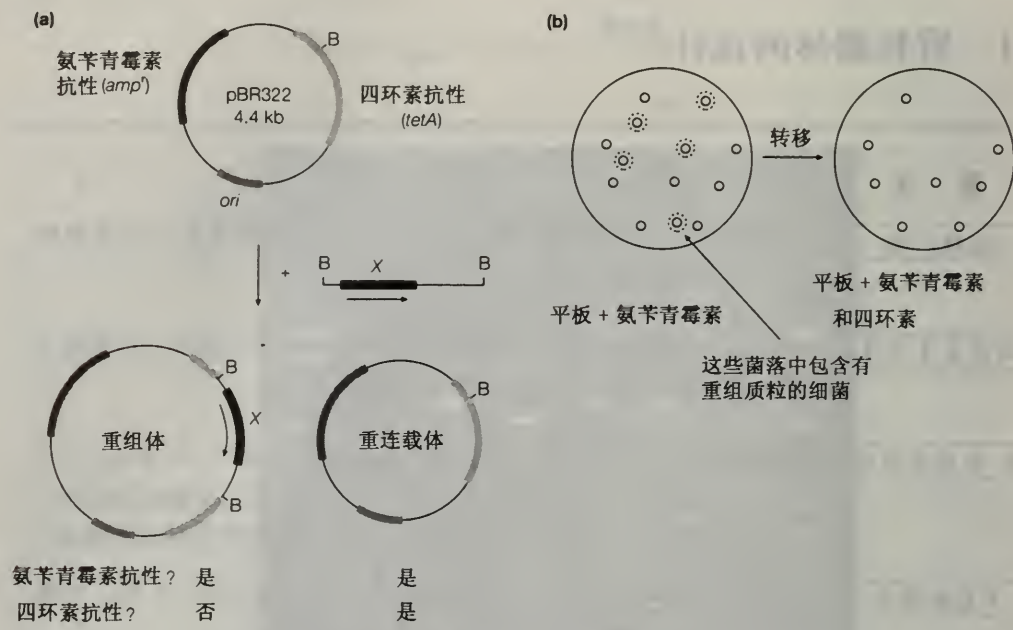


图 H1.1　(a)用抗性基因的插入失活法来筛选重组体；(b) 复制铺板

在转化及铺板步骤中（参见 G4），所有在氨苄青霉素平板上长成的菌落肯定都是含有完整的 amp^r 基因的质粒。自连接的载体也会表现出四环素抗性，而 *tetA* 基因中有目的片段插入的克隆将不产生有活性的 TetA 蛋白，会对四环素敏感。在最初筛选平板上加入四环素将会杀死我们想得到的重组菌落，因此需要运用一种更间接的筛选方法即**影印平板培养**（图 H1.1b）。在含氨苄青霉素的正常平板上长出的菌落，用一种有吸附性的垫子转移到另一个含四环素的平板上，能在该平板上生长的菌落很可能是空载体，而不能生长的菌落则极可能是重组体。后者可以从含氨苄青霉素的平板上挑取下来作进一步分析。

蓝-白颜色筛选

一种更巧妙的鉴定重组质粒的方法是蓝-白颜色筛选，该法可以在同一转化平板上完成筛选。这种方法也涉及基因的插入失活，正如名称所示，它利用一种蓝色化合物的形成作为指示剂。在这种情况下失活基因是 *lacZ*，该基因编码 β-半乳糖苷酶，受 *lac* 启动子的调控（参见 L1）。当 *E. coli* 菌株表达 **Lac 阻抑物**，则载体上的 *lacZ* 基因由**异丙基-β-D-硫代半乳糖苷**（IPTG）诱导表达（参见 L1），表达形成的酶可以利用底物 **5-溴-4-氯-3-吲哚-β-D-半乳糖苷**（X-gal）形成一种蓝色化合物。重组质粒中 *lacZ* 的插入失活将导致不能形成蓝色菌落。该方法具体操作是(图 H1.2）将转化细胞涂布于含氨苄青霉素（通常筛选转化子的方法）、IPTG 和 X-gal 的平板上，形成有蓝白两色的菌落混合群体。白色菌落未

表达β-半乳糖苷酶，因而可能含有插入的目的片段，而蓝色菌落则很可能是自环载体的克隆。

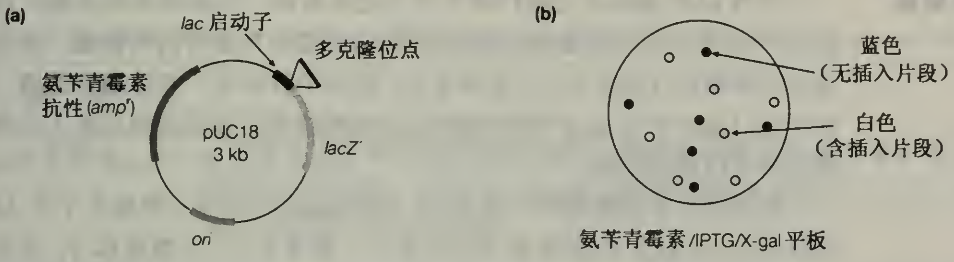


图 H1.2 (a)用于蓝白筛选的质粒载体；(b) 蓝白筛选实验中所形成的菌落

实际上这种方法中所用到的载体都有一个 *lacZ* 截短后的衍生物 ***lacZ′***，此基因编码β-半乳糖苷酶的 N 端的 **α 肽**。这种载体须转入一种含有表达β-半乳糖苷酶的 C 端部分的突变基因的宿主菌，表达的β-半乳糖苷酶的 C 端能与α肽互补产生有活性的酶。这样做的好处是缩短了质粒所载基因的长度，但又没有改变该法的原理。

多克隆位点

第一个利用蓝白颜色进行筛选的载体，同时也开辟了**多克隆位点**（multiple cloning site，MCS）。这些质粒即 pUC 系列，包含一个改造过的 *lacZ′*基因，在该基因编码区的第一部分有多个限制酶酶切位点（图 H1.2a 和图 H1.3）。这一区域被称为 MCS。目的 DNA 插入于这些位点中的任一个或任两个位点之间，都会使 *lacZ′*基因失活，并在适宜的平板上产生白色菌落。MCS 使在选择用于克隆的限制酶时具有一定的灵活性。

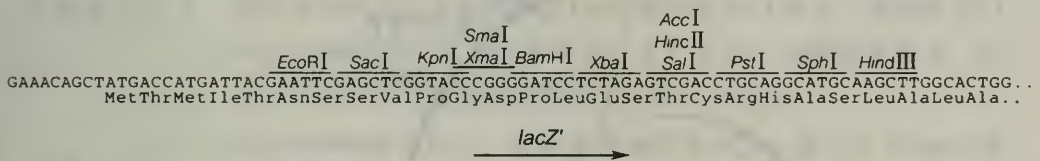


图 H1.3 在 *lacZ′*基因 5′端的多克隆位点

已克隆插入片段的转录

在 pUC 载体上与克隆片段的插入位点相邻处有一个启动子（*lac*），很易想象出这样的启动子可用来转录插入 DNA，无论是在体外产生可用作杂交探针的 RNA 转录物（参见 I3），还是表达插入片段内基因的蛋白产物（参见下文），都不成问题。

目前已构建了多种转录型载体，以便于在体外进行克隆片段的转录。例如 pGEM 系列载体中就含有来自**噬菌体 T7** 和 **SP6** 的启动子，两启动子间是 MCS（其本身是为 *lacZ′*蓝-白筛选而设计）。来自噬菌体的启动子只能被各自相应的噬菌体 RNA 聚合酶所识别（参见 G1.1），上述两启动子

的任一种均可用于体外转录插入片段的任一链。

表达载体

可用 pUC 载体在大肠杆菌中表达克隆基因。这就需要将克隆片段在 MCS 中定位以使目的基因的编码区与 *lacZ'* 基因具相同的**可读框**，并且是连续不间断的（参见 P1）。在诱导 *lac* 启动子转录后，产生**融合蛋白**，该融合蛋白是由来自 *lacZ'* 的 N 端的几个氨基酸及其紧连着的由插入片段所编码的蛋白序列组成。

使用强（活性极高的）启动子，例如 ***lacUV-5***［一种独立于环 AMP 受体蛋白（CRP）的突变型 *lac* 启动子，参见 L1］、λ 噬菌体 P_L 启动子（参见 R2）以及**噬菌体 T7** 启动子，已构建成很多种不同的载体。某些载体需要克隆片段提供核糖体结合位点（RBS）以及翻译起始密码子（参见 Q1），而另一些载体则设计成在表达蛋白的 N 端加上一段融合序列，以使重组蛋白可通过一特异结合步骤很方便地被纯化出来，如**组氨酸标签**（His-tag）即一串组氨酸残基便是这类序列的一个例子，该序列能够与带 Ni^{2+} 的色谱柱紧密结合（参见 B3）。另一些载体甚至可以将融合 N 末端用特异蛋白酶从蛋白质分子上切下来。

举个例子来说：T7 启动子系统能使许多蛋白在大肠杆菌中的过量表达（图 H1.4）。载体中含有一个 T7 启动子和一个大肠杆菌 RBS，紧接着是一个起始密码子、一个 MCS、最后是一个转录终止子（参见 K4）。经 IPTG 诱导后能在产生 T7 RNA 聚合酶的大肠杆菌菌株（一种 λ 溶原菌；参见 H2）中进行表达。该酶效率很高，且只对 T7 启动子进行转录，最佳状态下能使重组蛋白的表达量高达大肠杆菌菌体总蛋白量的 30%。

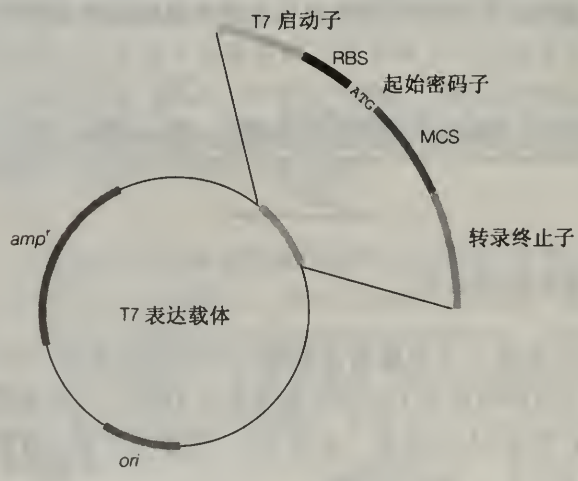


图 H1.4　用 T7 系统表达外源基因的质粒载体

H2 噬菌体载体

要 点

λ噬菌体	λ噬菌体对大肠杆菌的侵染及随后的细胞裂解过程可用来扩增克隆的 DNA 片段。线性 48.5kb 的λ基因组 DNA 中最大为 23kb 部分可由外源 DNA 所替换。
λ置换型载体	将目的 DNA 片段连接于为侵染所必须的基因的末端，形成重组λ噬菌体 DNA。
包装与侵染	包装的抽提液含有λ外壳蛋白及包装过程中所需酶，用它可将重组λ DNA 包装到噬菌体颗粒中，以实行对大肠杆菌的非常有效的侵染。
噬菌斑的形成	将λ侵染的细胞分布在未被侵染的大肠杆菌菌苔中会形成噬菌斑，是由于侵染和细胞裂解循环而抑制了细胞生长的区域。
λ溶原体	利用λ噬菌体的溶原生长期将克隆基因整合进大肠杆菌基因组，以使其长期表达。
M13 噬菌体载体	M13 噬菌体在大肠杆菌细胞内以双链环状形式复制，可以像质粒一样操作双链环状 M13 DNA，但产生的噬菌体颗粒内仅含有环状 ssDNA。
克隆到 M13	用标准质粒克隆方法将重组 DNA 整合到 M13 载体，重组的 M13 DNA 侵染敏感性大肠杆菌细胞，并形成生长缓慢的噬菌斑，然后从纯培养基中的噬菌体颗粒中分离出 ssDNA。
质粒-M13 杂合载体	在辅助噬菌体帮助下，对宿主细胞的侵染，可诱导含有 M13 复制起始点的质粒形成单链噬菌体颗粒。

相关主题 基因操作（G） 核酸测序（J2）
基因文库与筛选（I）

λ噬菌体　　侵染大肠杆菌细胞的λ噬菌体可被用作克隆载体。λ噬菌体的侵染过程将在主题 R2 中有详述。简而言之，噬菌体颗粒将其**线性 DNA** 注入细胞，进而 DNA 连成环状，其后该 DNA 被复制组装成噬菌体颗粒，经裂解细胞释放到胞外，造成细胞死亡（**裂解**期），或通过特异位点将其 DNA 整合到宿主基因组（参见 F4），而获得长期插入（**溶原**期）。

图 H2.1 显示 48.5kb 的 λ 基因组，在其末端有由 12bp 的黏末端构成的 *cos*（黏性）位点。*cos* 位点是不对称的，但从另一角度来看，*cos* 位点相当于较长的（16bp）限制性位点（图 H2.1 或参见 R2），可以导致 DNA 在细胞内自身环化。基因组中心区域的大部分对侵染裂解都是不必要的，可以被不相关的 DNA 序列所替换。不过对插入 λ 噬菌体颗粒的 DNA 的大小是有一定限制的，即 DNA 的总长度必须是原长度的 75%～105% 之间，即 37～52kb。考虑到必要区域，可克隆进 λ 噬菌体的 DNA 片段大约为 20kb（最大 23kb）左右，远远大于可插入质粒载体的片段。

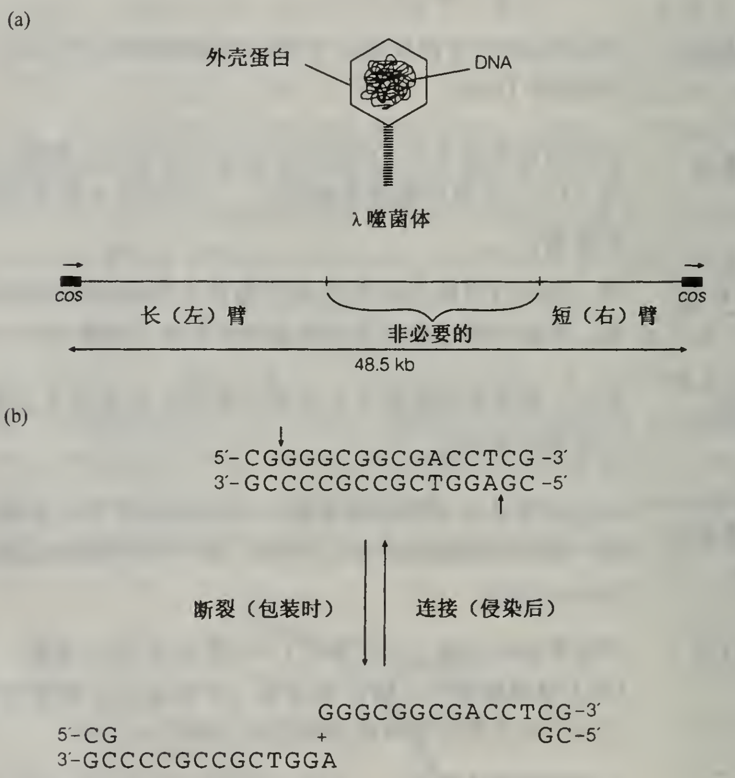


图 H2.1　(a) λ 噬菌体及其基因组；(b) λ 噬菌体的 *cos* 末端

λ 置换型载体

用 λ 噬菌体开发出了许多**置换型载体**，例如 **EMBL3** 和 **λDASH**（参见 I1）。图 H2.2 展示了利用这种载体进行克隆实验的一个代表性方案。用 *Bam*HⅠ酶解载体 DNA，回收长臂（19kb）与短臂（9kb）（图 H2.1）。同样用 *Bam*HⅠ或其他适合的酶酶解并回收目的片段（参见 G3 和 I1），同时可用碱性磷酸酶处理以防止片段间相互连接。λ 臂与目的片段在较高浓度下相互连接形成较长的线性产物（图 H2.2）。

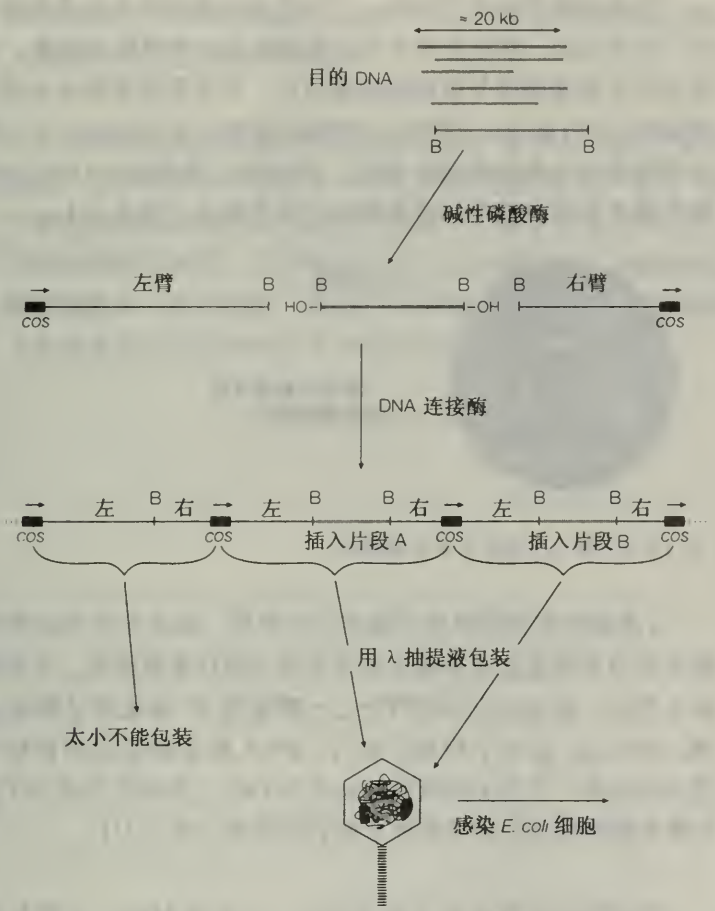


图 H2.2 用 λ 置换型载体进行的克隆实验

包装与侵染

虽然很纯的环状 λ DNA 或其衍生物可以与质粒相同的方法转化感受态细胞（参见 G4），但噬菌体颗粒的侵染性有其优点，特别是用于 DNA 文库的构建（参见 I），λ 噬菌体在体内复制产生多拷贝的 λ 基因组长线性分子。将这些多联体在 *cos* 位点处切开，产生单个 λ 基因组，再被包装进噬菌体颗粒。噬菌体外壳蛋白及噬菌体 DNA 加工酶的混合液（**包装抽提液**）在体外可将连接后的线性分子**包装**成噬菌体颗粒（图 H2.2）。包装抽提液可从感染不同**包装缺陷型**（突变）λ 噬菌体的两个菌株来制备，因而抽提液中的包装蛋白是丰富的。来自两个菌株的抽提混合液提供了包装所需的所有蛋白。这样形成的噬菌体颗粒可侵染正常的大肠杆菌细胞。不含插入片段的 λ 末端连接物，或远远小于或远远大于最佳长度 20kb 的插入片段的连接物对包装都是过小或过大，不能被包装，同时含有两条左臂或两条右臂的重组体也是不能生存的。侵染过程是非常有效的，且每毫克载体 DNA 可以产生出多至 10^9 个重组体。

噬菌斑的形成

先将高浓度未被侵染的、可培养形成连续菌苔的细胞涂布于琼脂平板上（参见 G4），然后再涂布来自包装反应的被侵染的细胞。在培养后的菌苔中由于反复感染与细胞的裂解循环，单个感染细胞处会形成空白区域即噬菌斑（图 H2.3）。这类似于细菌单菌落（参见 G4）。为了进一步研究可从噬菌斑中分离出噬菌体颗粒，再纯化出重组的 λ DNA，或者也可从用特异重组噬菌斑感染的培养物的上清中纯化（参见 J1）。

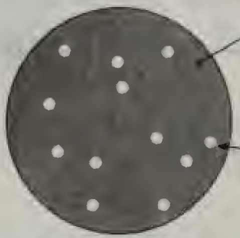


图 H2.3　经 λ 感染所形成的噬菌斑

λ 溶原体

λ 感染的溶原期也被克隆技术所利用。通过含有目的序列的 λ 载体的整合作用可将基因或外源序列整合进大肠杆菌基因组，这种重组基本上是永久性的。该方法应用的例子之一就是用 T7 系统来过量表达蛋白质的菌株。含有 *lac* 启动子控制下的 T7 RNA 聚合酶基因的菌株 BL21（DE3）及其衍生物，作为 λ 溶原菌被命名为 DE3。该基因可被 IPTG 诱导后，然后聚合酶将转录出表达载体上的目的基因（参见 H1）。

M13 噬菌体载体

所谓的丝状噬菌体（参见 R2），尤其是 **M13**，可作为大肠杆菌的载体。噬菌体颗粒含有 6.7kb 环状单链 DNA。在感染感受性 *E. coli* 细胞后合成互补链，形成双链环状 DNA，即复制型（RF），每个细胞中约含有 100 个拷贝。与 λ 噬菌体相比，M13 侵染的细胞并不裂解，而是继续缓慢生长，单链形式的 DNA 不断被包装，并从细胞中释放出去（每一个细胞多至 1000 个）。互补双链中相同的单链总是存在于噬菌体颗粒中。

作为载体，M13 的一个可利用的特性是其复制型（RF），可以精确地像质粒一样被纯化与操作，而相同的 DNA 也可以单链形式从培养基中的噬菌体颗粒分离获得。ssDNA 有许多种应用，包括 DNA 测序（参见 J2）和定点诱变（参见 J5）。

克隆到 M13

并不用 M13 作为克隆新的目的 DNA 的初级载体，而是当需要单链片段时，用标准质粒操作法将片段亚克隆进 M13 RF（参见 G）。用重组 DNA 转染大肠杆菌细胞，并涂布平板，会形成像 λ 噬菌体感染相同的噬菌斑，只不过噬菌斑是由于侵染细胞生长缓慢而不是裂解形成的。利用 MCS 和 *lacZ′*（参见 H1）的蓝白筛选已组进 M13 载体，如 M13mp18 和 19，它们极其相似，只是相对于 M13 复制起始点其 MCS 方向相反。复制

起始点决定了包装进噬菌体的链，所以可根据目的片段的哪条链需要被装入噬菌体来决定使用哪种载体。

质粒－M13 杂合载体

许多较小的质粒载体如 pBluescript，都已开发成组合有 M13 功能的载体。它们既含有质粒也含有 M13 的复制起始点，但并不拥有为噬菌体完整生活周期所需的各种基因。因而它们通常像真正的质粒一样扩增，具有质粒载体生长快、易操作的优点，但在需要时，这些载体与功能完整的**辅助噬菌体**一起共感染，诱导产生单链噬菌体颗粒。这里辅助噬菌体提供为单链复制与包装所需的各种基因产物。

H3 黏粒、YAC 与 BAC

要 点

大片段 DNA 的克隆	真核生物基因和真核生物基因组结构的分析需要比质粒和 λ 噬菌体更能容纳大片段 DNA 的载体。
黏粒载体	黏粒用 λ 包装系统将于 λ*cos* 位点结合的大片段 DNA 包装起来，在侵染大肠杆菌细胞后环化体能像质粒一样复制。某些黏粒载体有两个 *cos* 位点，酶切产生后两个 *cos* 末端，可与目的片段的两端连接，然后包装入 λ 颗粒。黏粒克隆 DNA 的容量约为 30～45kb。
YAC 载体	酵母人工染色体是将酵母染色体中为复制与分离所需序列与大片段目的 DNA 连接而构成的，可克隆外源片段的长度可大于 1Mb。酵母人工染色体（YAC）载体含有两个端粒序列（TEL）、一个着丝粒（CEN）、一个自主复制序列（ARS）以及能在酵母中用作筛选标记的基因。
用酿酒酵母的筛选	在酵母中筛选 YAC 或其他载体是利用与不能产生某种基本代谢的突变菌株的互补来进行的，即 YAC 载体上携带该突变基因的正常拷贝可以互补这一缺陷。
BAC 载体	细菌人工染色体是基于大肠杆菌 F1 因子的，在便于操作的大肠杆菌宿主中能克隆长达 350kb 的基因组 DNA，比 YAC 更稳定、易于操作。
相关主题	质粒载体的设计（H1） 噬菌体载体（H2） 基因组文库（I1）

大片段 DNA 的克隆

对于庞大基因组的生物（如人类 DNA 有 3×10^9bp），其基因组结构分析及基因鉴定很难使用质粒和 λ 噬菌体载体来进行（参见 G 和 H2），因为使用此类容量很小的载体来克隆基因就意味着构建一个 DNA 文库就需要数目巨大的克隆群才能覆盖整个基因组（参见 G1 和 H1）。此外，一些由于较大的内含子序列的真核生物基因也无法将整个基因放在一个克隆片段中。为解决上述这些问题，已开发出具有大 DNA 容量的载体。**脉冲场凝胶电泳**（PFGE）的出现使大片段 DNA 的分离、作图和分析成为可能。在 G3 中就会看到，传统的琼脂糖凝胶电泳有十分明显的缺点：DNA 大片段分不开，而是共迁移。这是因为核酸在多孔基质如琼脂糖凝胶中迁移

时，折叠（球状）型和伸展（线状）型分子交替出现。当 DNA 分子很大时，其球型体与基质网孔就不匹配了，即使在容易操作的低浓度(0.1%～0.3%）琼脂糖中，它们仍然是共迁移。如果电场是不连续的（脉冲的），则可以克服这个缺点。如果电场方向也是可变的，则能分离更大的 DNA 分子。每当电场改变时，DNA 分离重新调整其长轴方向，而较大的分子需要较长的定位时间。根据其电极数目和电场的变化的不同，已开发出多种 PFGE［例如倒转电场凝胶电泳（FIGE）和钳位均匀电场电泳，(CHEF)］。已经能分离长达 7Mb 的分子，现已可解决整个染色体 DNA 片段的分离问题，包括人工染色体的分离（参见下文）。

黏粒载体

黏粒载体的名称是因为该载体以质粒分子形式，且利用了 λ 噬菌体 *cos* 位点的特性而得名。体外将 DNA 包装入 λ 颗粒（参见 H2）仅要求 λ*cos* 位点在线性 DNA 上有正确的距离（37～52kb)。间插 DNA 可以是任

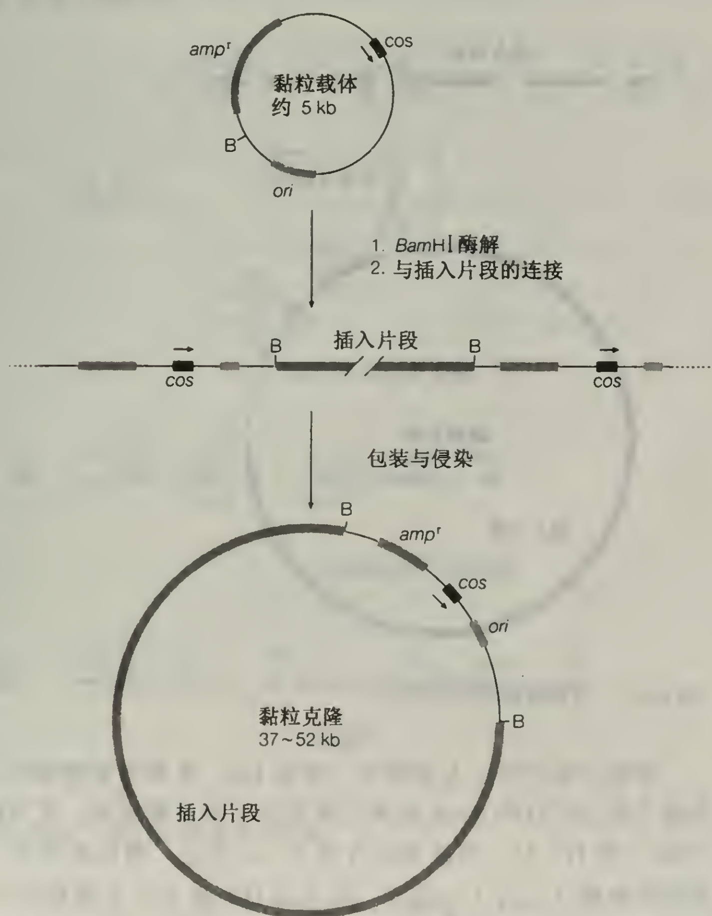


图 H3.1 一个黏粒克隆的形成

何序列的 DNA，例如并不需要含有任何 λ 基因。最简单的黏粒载体是一个正常的小质粒，包含质粒的复制起点（*ori*）、选择性标记以及一个 *cos* 位点和一个用于克隆的限制性位点（图 H3.1）。用限制酶降解，与目的 DNA 片段连接后，DNA 将被包装进 λ 噬菌体颗粒。感染后经 *cos* 位点的配对，DNA 重新环化，在氨苄青霉素选择压力下像正常质粒一样扩增。



图 H3.2　用黏粒载体进行克隆

实际克隆中对于 λ 噬菌体（参见 H2）常采取更精细的方法以防载体的多个拷贝或目的 DNA 的多个拷贝进入某个重组体。其中的一个例子是 C2XB（图 H3.2），该载体含有两个 *cos* 位点，两位点间有一个产生平末端的限制酶（*Sma* I）位点。用 *Bam*H I 和 *Sma* I 酶切后产生出两个 *cos* 末端，而后与用碱性磷酸酶处理（参见 G1 和 G4）过的目的 DNA 片段相连接。在所定条件下 *Sma* I 的平末端连接不能有效进行，最终能被包装的产物只是图中所示的产物，即产生出插入 30～45kb 片段的重组黏粒。

用黏粒载体所制备的克隆文库可以用主题 I3 中描述的方法进行筛选。经多年努力，整个大肠杆菌基因组已可被一套约 100 个的黏粒克隆所覆盖。

YAC 载体

弄清了真核生物染色体为稳定复制与分离所需的序列，至少在酿酒酵母中，是由相当小且确定的序列（参见 D3）构成的事实后，于是构建成了重组染色体（**酵母人工染色体**，YAC）。起初只是为了了解染色体的稳定性，后来就开发成能携带巨大克隆片段的载体。着丝粒、端粒及复制起点序列（参见 D3、E1 和 E3）已经被分离，并且组合到大肠杆菌的质粒上。典型 pYAC 载体的结构见图 H3.3。YAC 克隆的构建方法类似于黏粒，只不过载体片段的两端与目的 DNA 相连构成完整的染色体，再导入（转染）到酵母细胞中（参见 H4）。YAC 载体可容纳大于 Mb 的基因组 DNA 片段，因而可以用于克隆完整的人类基因，比如长 250kb 囊性纤维化基因。YAC 用来对巨大基因组的大规模结构进行作图时很有效，例如人类基因组计划。

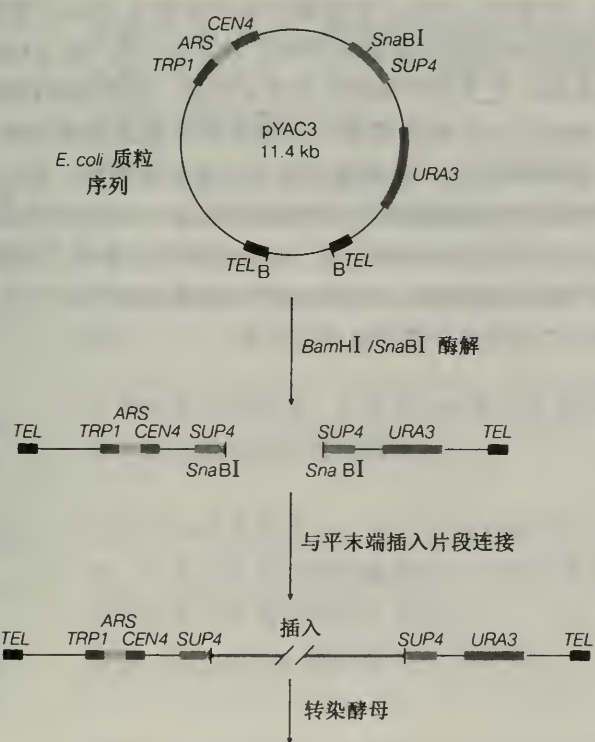


图 H3.3 用 YAC 载体进行克隆

pYAC3（图 H3.3）所携带的酵母序列分述如下：TEL 代表端粒 DNA 序列片段，该片段可以被酵母中的端粒酶延伸（参见 D3 和 E3）；CEN4 是酿酒酵母 4 号染色体的着丝粒序列，不同的是，即使着丝粒紧邻于人工染色体的一端，它也可以正确地分离子染色体；ARS（**自主复制序**

列）的功能是酵母复制起点（参见 E3）。*TRP1* 和 *URA3* 是酵母选择性标记（参见下文），分别位于两端，以确保只有正确重组的 YAC 才能在酵母细胞中生存；*SUP4*，在重组体中行插入失活功能（图 H3.3），是一个相关于红-白颜色测试的基因，类似于大肠杆菌中的蓝-白筛选（参见 H1）。

用酿酒酵母的筛选

酿酒酵母选择性标记并不像大肠杆菌的质粒那样能赋以对毒性物质产生抗性，而是能使酵母在缺乏某种特定养分的选择性培养基中生长。原理是**营养缺陷型**酵母突变株不能产生某种特定化合物。例如 *TRP1* 突变株，不能合成色氨酸，只有在添加了色氨酸的培养基中才能生长。携有完整的 *TRP1* 基因的 YAC（或其他载体，参见 H4）突变酵母株的转化可**互补**这一缺陷，结果是只有被转染的细胞才能在缺乏色氨酸的培养基上生长。

BAC 载体

BAC 载体即**细菌人工染色体**克服了在用 YAC 克隆非常大的基因组 DNA 片段时所遇到的问题。YAC 能容纳非常大的片段，这些片段经常形成非毗邻（非邻近）片段，在增殖中会丢失部分 DNA（即不稳定）。BAC 能容纳长达 300～350kb 的插入序列，比 YAC 短，但 BAC 的优点是：稳定、容易转化，在大肠杆菌宿主中生长迅速，应用标准质粒小量制备技术纯化也较简单。BAC 载体是基于大肠杆菌天然的染色体外 F 因子，并编码自己的 DNA 聚合酶，在细胞中仅有 1 或 2 个拷贝。BAC 载体整合了 F 因子复制和保持所需基因以及 1 个选择标记、一个在稀有限制酶位点侧翼的克隆位点和其他特殊切割位点。特殊切割位点确保了克隆能在载体区域内线性化，而不需要在插入片段内进行酶切。BAC 比 YAC 更受使用者的青睐，现已广泛应用于基因组作图计划。

H4 真核生物载体

要 点

克隆到真核生物	许多基因克隆的应用都需要将基因转入真核细胞并获得表达，不论是暂时性还是永久性的。
转染真核细胞	将 DNA 转染真核细胞也许需要消化细胞壁（酵母及植物）或是将 DNA 沉集到细胞表面（动物细胞）。电穿孔可促进 DNA 的吸收。微注射或是固体粒子轰击也是可行的。
穿梭载体	许多真核载体含有细菌质粒序列，从而能在大肠杆菌宿主中构建并检测。它们可以在两种以上宿主间穿梭。
酵母附加型质粒	利用天然酵母的 2μ 质粒的复制起点及选择性标记如 *LEU2* 开发出酵母载体系列（YEp)。这些载体可像质粒一样复制，而且还可被整合进染色体 DNA。
根癌农杆菌 Ti 质粒	能够侵染某些植物并将其部分 Ti 质粒整合到植物基因组中的根癌农杆菌已被用来成功将外源基因转入多种植物。
杆状病毒	杆状病毒是一种昆虫病毒，用于在昆虫细胞培养中过量表达动物蛋白。
哺乳动物病毒载体	许多哺乳动物病毒，包括 SV40 及反转录病毒，已被用作哺乳动物培养细胞的基因操作载体。
直接基因转移	不用特殊的真核载体，也可直接将基因导入植物或动物培养细胞。携有真核基因的细菌质粒可以暂时保留在细胞中而不复制，也可低效重组整合进宿主基因组。
相关主题	克隆 DNA 的分析与应用（J）

克隆到真核生物

采取以大肠杆菌为宿主的基因克隆技术已从真核生物分离出了许多真核基因及其调控序列，并对其进行分析。然而遗传工程的许多应用（参见 J），如从大量生产真核蛋白到新植物工程及基因治疗，都需要能在各种真核物种中表达外源基因的载体。针对各种宿主所设计的这类载体的具体例子将在本主题中讨论。

转染真核细胞

将 DNA 转入真核细胞（**转染**）比转化细菌困难得多，同时转染的效率也低得多。例如对于酵母及植物细胞，通常要用降解酶将细胞壁消化掉以产生脆性**原生质体**，从而能较容易地吸收 DNA。一旦去掉降解酶，细胞壁仍能重新合成。相比之下，动物培养细胞没有细胞壁，可以较低效率吸收与**磷酸钙**一起沉集于细胞表面的 DNA。用高电压处理细胞可以提高转染效率，因为高电压可以在细胞膜上打开临时的孔道，该过程被称为**电穿孔**（electroporation）。

也可采用直接的物理方法如微注射或基因枪等。用极细的玻璃毛细管进行**显微注射**，直接将 DNA 注入到培养中的单个动物或植物细胞的细胞质甚至细胞核中。此外还可用**基因枪**等把包被着 DNA 的金属微粒打入靶细胞。

穿梭载体

大多数真核细胞中使用的载体都被构建成**穿梭载体**。这就意味着这些载体含有在大肠杆菌或目的宿主细胞中进行复制与筛选所需的序列（*ori*，amp^r）。这就使得携带目的插入片段的载体在转入适宜真核细胞前可以采用完备的大肠杆菌操作方法来进行构建及综合检测。

酵母附加型质粒

在酿酒酵母中用于基因克隆与表达的载体是基于 **2μ 质粒**而设计的。这个质粒是根据其 DNA 的长度而命名的，相当于约为 6kb 长。2μ 质粒含有一个复制起点和两个涉及复制的基因，还编码一个位点特异性的重组蛋白 FLP，该蛋白与 λ 噬菌体整合酶 Int 同源（参见 F4），Int 可以转换部分 2μ 序列。基于 2μ 的质粒载体（图 H4.1）被称为**酵母附加型质粒**（YEp），通常包含 2μ 的复制起点、大肠杆菌穿梭序列和一个可作为选择性标记的酵母基因（参见 H3），如相关于亮氨酸生物合成的 *LEU*2 基因。虽然这些载体通常像质粒一样地复制，但 YEp 也可以通过与选择基因的缺陷型基因组拷贝进行同源重组（参见 F4）的方式整合进酵母染色体（图 H4.1）。

根癌农杆菌 Ti 质粒

植物细胞中不含有能用作克隆载体的天然质粒。一种根癌农杆菌主要是侵染双子叶植物（番茄、烟草、豌豆等），最近显示的结果也能侵染单子叶植物如水稻。根癌农杆菌携有 200 kb 的 **Ti 质粒**（Ti，tumor inducing，诱导癌肿），侵染时 Ti 质粒的部分 **T-DNA** 会整合进植物染色体 DNA（图 H4.2），结果在 T-DNA 基因指导下造成植物细胞的失控生长，形成冠瘿瘤或根瘤。

在 T-DNA 区插入目的基因后重组 Ti 质粒就能把基因整合到植物 DNA 并且有可能表达。在实践中，整个程序需要精心设计。Ti 质粒的很大很难对其操作。研究发现 T-DNA 和 Ti 质粒的其余部分若在同一细菌细胞中的不同分子上，仍能发生整合。因此可在标准的大肠杆菌质粒上构建重组的 T-DNA，然后转化到根癌农杆菌细胞中，该农杆菌携有无T-DNA、

改造过的 Ti 质粒。进一步的改良是从 T-DNA 中删除冠瘿形成基因，这就是所谓**卸甲** T-DNA 穿梭质粒，用于整合克隆基因。如果宿主细胞在培养基中生长，则可从转化细胞再生出完整的重组植株（图 H4.3）。

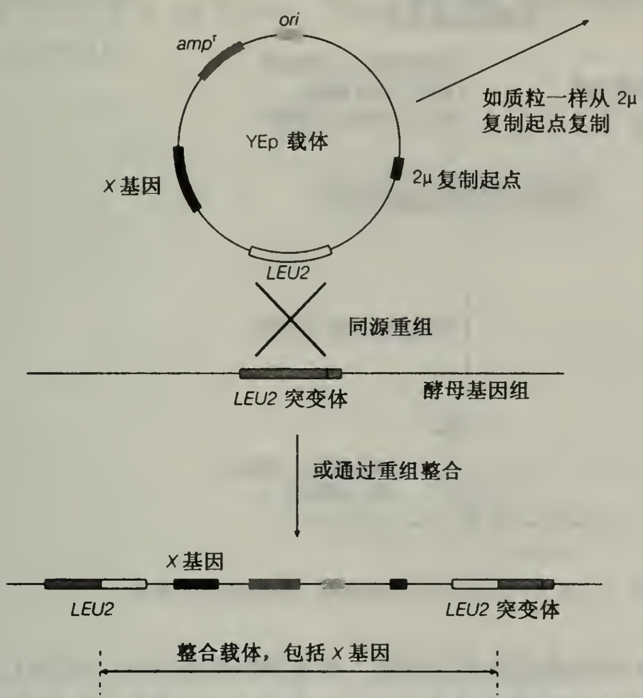


图 H4.1 利用基于 2μ 复制起始点的酵母附加型质粒进行克隆

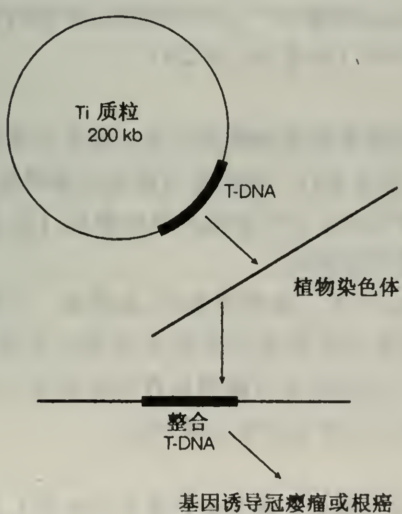


图 H4.2 *A. tumefaciens* Ti 质粒的作用方式

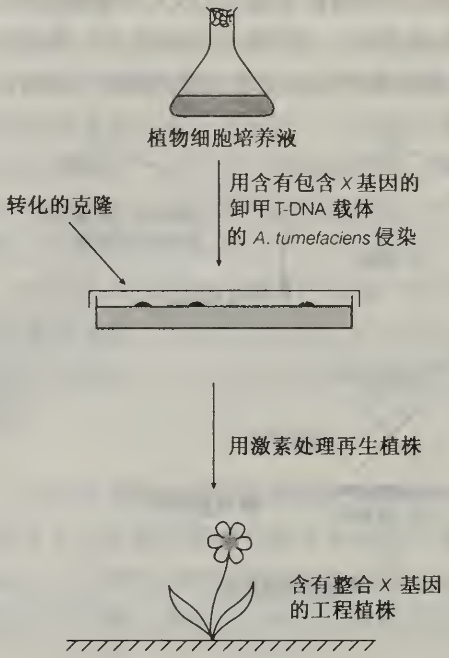


图 H4.3 用 *A. tumefaciens* 构建新工程植株的示意图

杆状病毒

杆状病毒侵染昆虫细胞。病毒基因组编码的主要蛋白之一是在感染细胞核内大量聚集多角体蛋白，这是因为该基因有一个极强启动子。该启动子同样可以驱动重组进杆状病毒基因组的外源基因的过量表达，从而经培养被感染的昆虫细胞来大量生产目的蛋白。这一方法已被广泛用于动物蛋白的大规模生产，因为昆虫细胞可产生许多翻译后修饰的动物蛋白，而细菌表达系统则做不到这一点（参见 H1 和 Q4）。

哺乳动物病毒载体

将基因转入哺乳动物细胞所用的载体也是以病毒为基础来构建的。第一个被利用的是 **SV40**（参见 R4），该病毒可感染多种哺乳动物。SV40 基因组只有 5.2kb，经历着类似于 λ 噬菌体的包装限制（参见 H2），所以利用 SV40 来转移大片段是有困难的。

反转录病毒（参见 R4）有一条单链 RNA 基因组，在感染后可拷贝成 dsDNA。该 DNA 可经类转位机制稳定地整合进宿主基因组（参见 F4）。反转录病毒有强启动子，所以被视为**基因治疗**中的载体（参见 J6），同时外源 DNA 也能以稳定的方式整合进宿主基因组。

直接基因转移

严格说来，可以不通过载体将基因暂时或永久地导入真核生物的培养细胞。例如细菌质粒上的真核基因可以在导入某一细胞系后瞬时表达其产物，即使在此类细胞中该质粒无法复制。此外通过转染或微注射导入的 DNA 也可以稳定地整合进细胞的染色体 DNA。在动物细胞中进行这种操

作通常要求外源 DNA 与基因组 DNA 间有显著的序列相似性（类似于图 H4.1），但对于植物细胞任何超螺旋质粒都可以随机整合进基因组，不过其具体过程尚不清楚。这类稳定转染的细胞可像细菌转化实验一样用抗药基因进行筛选（参见 G4），并且能够在多次细胞分裂中连续表达外源基因编码的蛋白。

（叶　彬 译　王薛林 校）

I1 基因组文库

要 点

具代表性的基因文库	由基因组 DNA 所制成的基因文库称为基因组文库，而由互补 DNA 所制成的称为 cDNA 文库，两者统称为基因文库。cDNA 文库不包括不能转录的核基因序列（重复序列等）。基因文库的好坏要看该文库是否覆盖起始材料的全部基因或序列而未因克隆操作丧失其中的基因或某些序列。
文库大小	基因文库必须包含一定数量的重组体，以便具有包含任何特定序列的可能性。如果知道基因组大小以及插入载体的片段的平均大小，则可以计算出所需重组体的数量。
基因组 DNA	为了构建文库，通常用蛋白酶降解及分相抽提制备基因组 DNA，要用物理剪切或限制性酶解来进行随机分割，以形成适合于所用载体大小范围的片段。常用一组限制酶对 DNA 进行部分降解。
载 体	质粒、λ 噬菌体、黏粒、BAC 或酵母人工染色体等载体都可被用来构建基因组文库，选用那种载体取决于基因组的大小。这些载体所能插入的片段大小的上限分别依次为 10kb、23kb、45kb、350kb 和 1000kb。用 T4 DNA 连接酶将基因组 DNA 片段与制备好的载体分子相连接。
相关主题	质粒载体的设计（H1）　　黏粒、YAC 与 BAC（H3） 噬菌体载体（H2）　　mRNA 加工、hnRNP 和 snRNP（O3）

具代表性的基因文库　　基因文库是来自某生物的不同 DNA 序列的总集，这些 DNA 序列都已被克隆进了载体以便于纯化、贮存与分析。依据所用 DNA 的来源，能构建的基因文库主要有两种类型。如果 DNA 来自基因组 DNA，该文库被称为**基因组文库**；如果 DNA 是 mRNA 群体的拷贝即 cDNA，则被称为 **cDNA 文库**。构建基因文库时，一个最重要的指标就是它能在多大程度上代表起始材料，即它是否能覆盖所有原初序列（**代表性文库**）？如果某些序列如缺少限制性酶切位点的重复序列（参见 D4）未被克隆，这样的文库就不具代表性。同样如果文库中未能含有足量的克隆，则极有可能会丢失某些基因。富含某种序列的 cDNA 文库（参见 I2）显然会缺少其他一些基因，但若正确制备并扩增，也可以成为富含 mRNA 起始材料的具代表性的文库。

文库大小

以最大可能可获得某一特定序列的基因文库所必须包含的重组体数（N）是可以计算的。公式是

$$N=\frac{\ln\ (1-P)}{\ln\ (1-f)}$$

其中 P 代表给定概率，f 是插入片段大小占总基因组大小的比例。例如给定概率为 0.99，插入片段大小为 20kb 的情况下，所需重组体的数对于 *E. coli* 基因组（4.6×10^6bp）与人类基因组（3×10^9bp）分别是

$$N_{大肠杆菌}=\frac{\ln(1-0.99)}{\ln[1-(2\times10^4/4.6\times10^6)]}=1.1\times10^3$$

$$N_{人类}=\frac{\ln(1-0.99)}{\ln[1-(2\times10^4/3\times10^9)]}=6.9\times10^5$$

这些数值可以解释为什么用插入大小仅有 5～10kb 的质粒就可以构建成很好的原核生物的基因组文库的原因，因为只需要几千个重组体就够了。对于较大的基因组，较大的插入片段可以使所需重组体数量较少，这正是开发黏粒以及 YAC 载体（参见 I3）的原因。此外用 λ 噬菌体作为克隆载体及其 λ 的高效率包装使之也不失为构建基因组文库时一个好选择。

基因组 DNA

为了构建具代表性的基因文库，须将纯化后的基因组 DNA 随机切割成适合克隆进所用载体的大小合适的片段。细胞分级分离会减少器官 DNA（线粒体、叶绿体）的污染。因此从真核生物纯化基因组 DNA 通常按照先制备细胞核，再用**蛋白酶降解**和**分相抽提**（酚/氯仿）除去蛋白质、脂类及其他杂质大分子的流程进行；而从原核细胞可以直接抽提 DNA。用该法制备的基因组 DNA 由染色体断裂而形成的数百 kb 大小的长片段构成。随机地将上述 DNA 切割成小片段的两种基本方法即**物理剪切与限制性酶解**。用吸管、振荡或**超声**（参见 C2）等物理剪切都可以非常随机地进一步将 DNA 切成小片段。方法的选择及实施的时间长短取决于所用载体对片段大小的要求。由于是对 DNA 双链的同时剪切所产生的末端，可能是平末端，即使一些不是平末端，可用 Klenow 聚合酶（参见 G1.1）补平。该 DNA聚合酶可在 dNTP 存在下将 DNA 分子凹进的 3′端补平。

因为限制性酶切位点的非随机分布，使用限制性内切核酸酶（参见 G3）酶解基因组 DNA 较易产生非随机片段。为了获得 15～25kb 或更大的基因组 DNA 片段（适于 λ 或黏粒载体的片段大小），需要进行**部分酶解**。通过使用较少量的限制性酶，使 DNA 上每一个限制酶识别序列不全部被酶解，结果可产生比完全酶解的情况下更长的片段。一种常用的酶是 *Sau*3A（识别序列 5′-/GATC-3′，/表示酶切位点），因为该酶解产生的黏性末端能与 *Bam*HⅠ（5′-G/GATCC-3′）酶解载体所产生的末端是相同的。对限制酶的选择须考虑到所形成的末端类型（黏末端或平末端），是否这些末端能与酶解后的载体直接相连，以及是否酶的作用会被 DNA

碱基修饰作用（如哺乳动物的 **CpG 甲基化**，参见 D3）所抑制。酶解时间及其用量应依据所产生的目的片段大小而变化（即能有效克隆进所用载体的片段大小）。合适的片段可用琼脂糖凝胶（参见 G3）或蔗糖梯度纯化回收。

载体

对于基因组较小的生物如 *E. coli*，可以用质粒载体来构建基因组文库（参见 H1），因为只需 5000 个克隆（平均插入片段 5kb）就可以高于 99%的概率覆盖整个基因组（4.6×10^6bp）。构建较大基因组的生物的文库时，多数是采用 **λ 噬菌体**、**黏粒**、**细菌人工染色体**（**BAC**）或**酵母人工染色体**（**YAC**）来进行。这些载体可以容纳的插入片段依次分别为 23，45，350 及 1000kb，因此覆盖整个基因组所需的重组体比使用质粒载体所需重组体要少得多。最常用的基因组克隆载体是 **λ 替换载体**（EMBL3，λ DASH），其典型克隆方案见 H2 中的图 I1.2。λ 载体 DNA 须被限制酶解切割成两个 λ 末端片段或称 λ 臂，以便将基因组 DNA 连于其间。原初载体在酶解后须通过梯度将中间片段（填充片段）与 λ 臂相分离而得以纯化，而现在的新型 λ 载体，则可用多种限制酶来进行酶解使填充片段不再被克隆。λ 臂制备好以后，按上述方法制备的基因组 DNA 可以用 **T4 DNA 连接酶**与 λ 臂相连（参见 G1 和 G4）。连接后的重组分子即可进行**包装**并**扩增**而建成文库（参见 H2）。

I2 cDNA 文库

要 点

mRNA 分离、纯化与分级分离	用与寡聚(dT)共价结合的磁珠很容易从真核细胞的裂解液中提取出 mRNA。mRNA 经其 poly(A)尾部与寡聚(dT)相结合，而从溶液中分离出来。制备的 mRNA 的质量可以用麦胚抽提液或网织红细胞裂解液进行翻译后，再由聚丙烯酰胺凝胶电泳检测翻译产物的方法进行检验；其质量还可用凝胶电泳进行直接检测，并通过回收凝胶泳道中特定区段按分子质量对 mRNA 进行分级分离。用杂交法可将特异序列从 mRNA 中分离出来。
cDNA 的合成	第一条链的合成是通过向引物，通常为寡聚(dT)的 3′端添加脱氧核糖核苷酸，用反转录酶来合成 mRNA 的 cDNA 拷贝。合成反应可由示踪标记进行监测。用末端转移酶对 cDNA 第一链的 3′端进行加尾可以很容易合成出全长的第二链。反转录酶或 Klenow 酶都可与同聚物尾部[例如，poly(dC)]退火配对来延伸引物[例如，寡聚(dG)]来合成第二链 cDNA。
cDNA 末端的处理	为避免 cDNA 与载体间的平末端连接，通常用单链特异性核酸酶对 cDNA 末端进行处理，再用 Klenow 酶补平后加上接头。为避免当添加的接头被限制酶降解时 cDNA 被酶解，可以在添加前对 cDNA 进行甲基化处理。连接分子也可以用作接头的替代物。
与载体的连接	载体通常被碱性磷酸酶去磷酸化以防止自连接，结果会促进重组分子的形成。质粒或噬菌体载体可用于构建 cDNA 文库，而噬菌体 λ gt11 最适合于用作构建表达文库的载体。

相关主题

DNA 克隆概述（G1） 基因组文库（I1）
质粒载体的设计（H1） 筛选流程（I3）
噬菌体载体（H2） mRNA 加工、hnRNP 和 snRNP（O3）

mRNA 分离、纯化与分级分离

通常不用原核 mRNA 来构建 cDNA 文库，因为原核 mRNA 通常不稳定；相比之下则较易制备基因组文库，且可包含所有基因组序列。从真核 mRNA 构建 cDNA 是非常有用的，因为 cDNA 不含内含子序列，可被用来在大肠杆菌中表达编码蛋白。由于 cDNA 从 mRNA 合成而来，cDNA 代表了基因组中可转录的部分（即为基因而不是非转录 DNA）。而且每类细胞或组织只表达一套特定的基因（但在受刺激或分化时可能有所改变），因此从特定组织中制备的 mRNA 常常会富含某种特异序列，如红细胞中

的珠蛋白 mRNA。所以对作为起始材料（组织）的选择要考虑到有利于 cDNA 克隆。在真核生物中，大部分 mRNA 都是聚腺苷酸化的（参见 O3），其约 200 个腺苷酸残基构成的 3′尾部为分离真核 mRNA 提供了有效的方法。**寡聚（dT）**可以与 poly（A）尾相结合，被用于回收 mRNA。传统方法是将提取的总 RNA（用酚-氯仿抽提细胞裂解液）通过**寡聚（dT）纤维素**柱来制备。但为了能将核酸酶造成的损伤降至最低，现在常采用一种更迅速的方法即直接将结合着寡聚（dT）的磁珠加入到细胞裂解液中，再用强磁铁将**磁珠**吸出再洗出 mRNA。某些情况下，裂解细胞后用**蔗糖梯度**（参见 A2）来制备 mRNA-核糖体复合物（多聚体，参见 Q1）可以作为提取 mRNA 的替换途径。

在用制备的 mRNA 进行 cDNA 克隆前，最好检测一下 mRNA 是否已被降解。检测可通过**翻译** mRNA 或**凝胶电泳**分析来进行。**无细胞翻译系统**如**麦胚抽提液**或**兔网织红细胞裂解液**通常被用来检测制备的 mRNA 的完整性。也可以将制备的 mRNA 显微注射到细胞内检测其翻译产物。用琼脂糖或聚丙烯酰胺凝胶来分析 mRNA 也是很常用的方法。制备的 mRNA在凝胶上常常会形成大小在 0.5kb 到 10kb 以上的分子范围内的不清晰的成片条带。通常两种最大的 rRNA 会污染 mRNA 制备物，由于其大量存在会在 mRNA 的成片条带中形成清晰条带。

有时在克隆 cDNA 之前分级分离或富集 mRNA 是很有效的，特别是当要克隆某一特定基因而不是构建完整 cDNA 文库时。分级分离是基于 mRNA 的大小进行的，按 mRNA 的不同大小分别从琼脂糖凝胶中回收。富集常通过**杂交**进行。例如当想要构建一个由激素、细胞因子或药物处理细胞而诱导产生的全部 mRNA 的 cDNA 文库时，须同时用诱导与未诱导的细胞来制备 mRNA。从未诱导细胞 mRNA 制备的第一链 cDNA（参见下文）可用来与诱导细胞的 mRNA 杂交。两种 mRNA 制备物中的共有序列会杂交形成双链体，而新诱导产生的 mRNA 则不会被杂交。不形成双链体的 mRNA 则可以被分离出来（如使用磁珠）用于构建 cDNA 文库。这样的文库可称为**扣除 cDNA 文库**。

cDNA 的合成

图 I2.1 显示出制备 cDNA 的方案。第一链合成时，**反转录酶**（参见 G1，表 I2.1）需要一个**引物**用以延伸以合成 mRNA 模板的拷贝。为了合成完整的 cDNA 通常引物为寡聚（dT），但若想获得**随机引物 cDNA**，引物也可以是六聚核苷酸所有可能组合的随机混合物。合成需要加入四种 dNTP。如果加入少量某一放射性标记 dNTP，则可用凝胶分析来测定 cDNA合成的效率。这称为**示踪标记**。通过向延伸中的引物的 3′端添加互补的核苷酸来拷贝模板。然而不幸的是，反转录酶容易从模板上脱落下来，尤其是碰到二级结构集中的区域，因此有时很难一步获得完整的（全长）cDNA 分子。

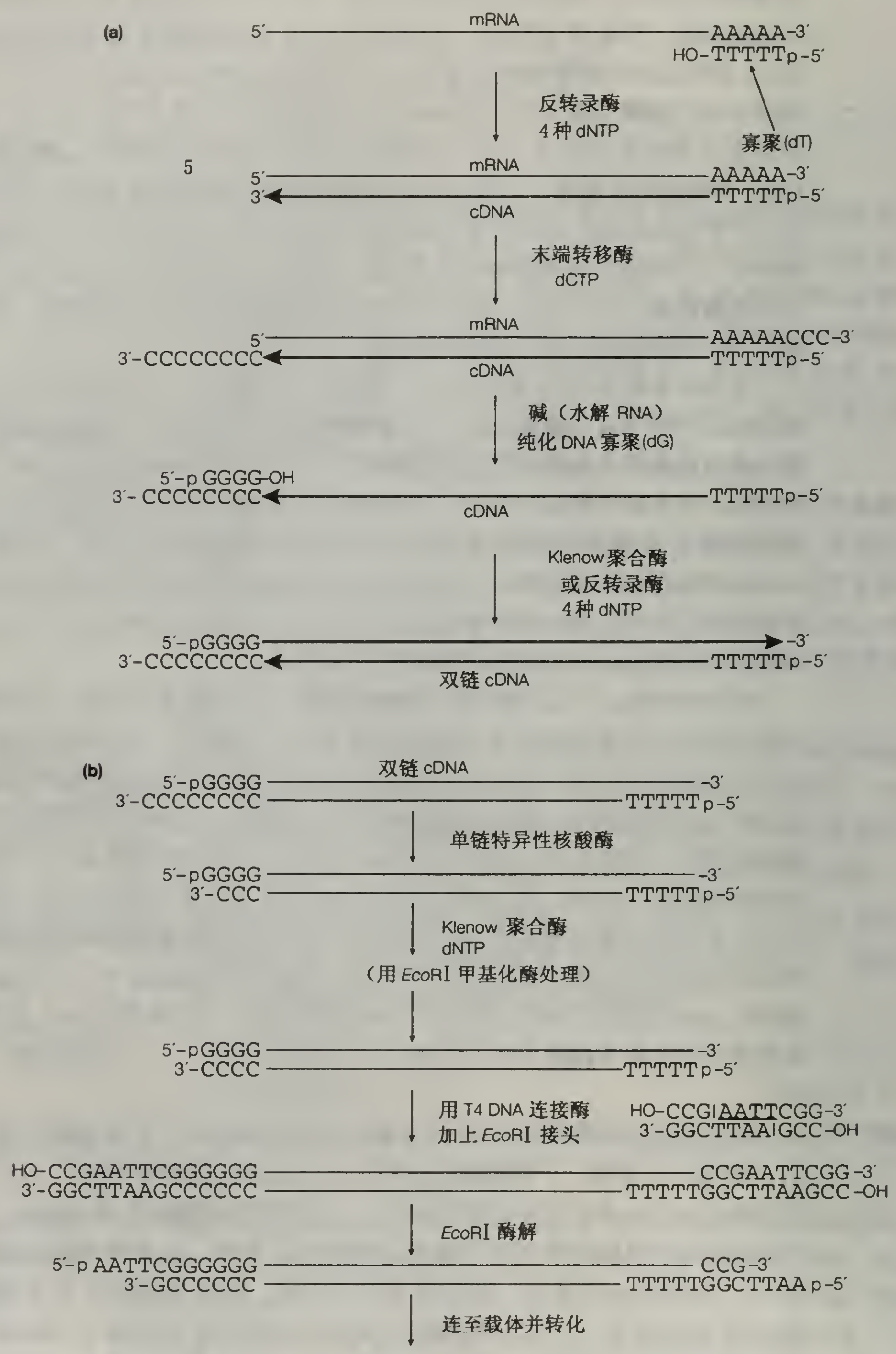


图 I2.1 cDNA 克隆过程。(a) 第一与第二链的合成；(b) 双链 cDNA 末端的形成及其接头添加

第二链的合成也需要引物。虽然可供选择的引物很多，但获得全长

cDNA 的最佳方案也许是通过对第一链的 3′端“加尾”，用与之互补的引物来合成第二链。**末端转移酶**（参见 G1，表 I2.1）可无需模板向单链或双链核酸的 3′端添加核苷酸。如果只提供一种 dNTP，如 dCTP（图 I2.1），则将在 mRNA-cDNA 杂合双链的 3′端加上由 C 残基的**同聚尾**。在用碱降解 mRNA 链后（参见 C2），可以用寡聚（dG）作引物，在反转录酶或 *E. coli* DNA 聚合酶 Ⅰ 的 **Klenow 片段**（参见 G1，表 I2.1）的催化下合成第二链。这一反应的产物是双链 cDNA，但分子的末端可能还不适合用于克隆。第一链的 3′端可能会超出第二链的 5′端，具体要看 poly (dC) 尾与寡聚（dG）引物的相对长度。因此有必要对 cDNA 的末端进行加工。

cDNA 末端的处理

大片段的**平末端连接**是效率很低的，所以为了避免用平末端与载体连接，而对双链 cDNA 的末端进行加工是很值得的。通常是添加特异核酸**接头**以形成适合于克隆的黏性末端。因为可以过量添加接头，cDNA 与接头的平末端连接反应进行得很顺利。具体讲添加接头第一步是用单链特异性核酸酶（S1 或绿豆核酸酶）处理 cDNA 切除掉凸出的 3′端。然后再用 DNA 聚合酶 Ⅰ 的 Klenow 片段和 dNTP 来补平 3′端。如果接头含有cDNA 上也具有的限制性酶切位点（如 *Eco*R Ⅰ，图 I2.1b），则 cDNA 应在添加接头前先用 ***Eco*R Ⅰ 甲基化酶**进行甲基化（参见 G3）。这用以保证限制性内切核酸酶不会在 cDNA 内部进行切割。随后可以用 T4 DNA 连接酶（参见 G1.1）将接头与双链 cDNA 平末端连接。如果接头未被磷酸化，则每个 5′-磷酸化的 cDNA 末端上可以连接一个接头，否则会连接上多个接头。最终用 *Eco*R Ⅰ 进行的限制性酶解将会产生可以与载体进行连接的黏性末端。有许多方法可以用来制备 cDNA 末端，其中包括用末端转移酶向 cDNA 末端加上与克隆载体末端互补的“尾部”，或是使用**衔接分子**来行使黏末端的功能，以免进行 cDNA 的甲基化反应。

与载体的连接

任何带有 *Eco*R Ⅰ 位点的载体都可用于图 I2.1 中的 cDNA 克隆。由于 cDNA 相对较短（0.5～10kb），所以常采用质粒载体；而为了大量克隆，尤其是为了构建表达 cDNA 文库，常常选用 λ 噬菌体载体（参见 H2）。连接前常常用碱性磷酸酶将载体去磷酸化，避免连接中载体片段自身连接，以保证产生的是载体与 cDNA 连接形成的重组体。载体 λgt11 有一个位于其 *lacZ* 基因的 C 端附近的 *Eco*R Ⅰ 位点，从而能将 cDNA 表达成 β-半乳糖苷酶融合蛋白的一部分。这将有助于文库的筛选（参见 I3）。常用 T4 DNA 连接酶连接载体与 cDNA，重组分子可以通过包装（参见 H2）或转化（参见 G2）来构建成 cDNA 文库。

I3 筛选流程

要 点

筛选	从一个基因文库分离出某个特定克隆的筛选通常需用一个核酸探针进行杂交，该探针可与其互补序列相结合从而检测出目的克隆。
菌落及噬菌斑杂交	将平皿上菌落或噬菌斑位置复制于膜的表面后，将其置于含有标记探针的溶液中进行保温。杂交并洗膜后，结合了探针的位置就可以被检出。将结合有标记的克隆或噬菌斑稀释后，再涂板进行下一轮筛选直至获得单克隆为止。
表达筛选	cDNA 表达文库中的 1/6 的克隆将以载体编码的 β-半乳糖苷酶的融合蛋白形式表达它们所编码的多肽。目的蛋白的抗体可被用来按类似于噬菌斑杂交的流程来筛选文库从而获得特定克隆。
杂交扣留与释放	cDNA 克隆可与 mRNA 样品进行杂交以阻止或抑制某些 mRNA 随后的翻译。另一方面通过杂交，那些与 cDNA 克隆杂交的 mRNA 也可以得到纯化，从 cDNA 库中释放出来，被翻译后用于编码多肽的鉴定。
染色体步移	这是为了获得重叠克隆群而对基因组文库进行的重复筛选，因而可构成覆盖一个染色体的某部分所有克隆的总汇。染色体步移还涉及用插入片段的末端作为探针进行连续筛选。染色体跳查是一个相似的过程。
相关主题	质粒载体的设计（H1） 克隆的鉴定（J1） 噬菌体载体（H2） mRNA 加工、hnRNP 和 snRNP（O3）

筛选

从基因文库的大量克隆中鉴定出某个含有目的基因的特定克隆的过程称为**筛选**。筛选中如要将一段 cDNA 片段或一段相关序列用作**核酸探针**则需要了解该基因或其产物的某些信息。如果能够获得足够多的蛋白质产物则可以测定出一段氨基酸序列（参见 B2），那么就可以此信息派生出可编码该氨基酸序列的所有可能 DNA 序列的混合物。该 DNA 序列信息可用来制作用于**杂交**筛选文库的核酸探针。制作用于文库筛选的 DNA 探针最普通方法之一是用聚合酶链反应（PCR，参见 J3），此法制作的探针叫做 **PCR 探针**。较短的核酸探针（**寡聚核苷酸**）很容易用自动化学合成仪合成，这些合成物可以直接用于杂交或制作更长 PCR 探针用于杂交。如果能设计一对引物，PCR 也可用作一种筛选文库的技术，因为一种 PCR 产

物仅当该序列（也就是克隆）存在时才将被检测到，并且该克隆库可以连续地亚克隆，直到分离出一个单一的阳性克隆。如果已由某蛋白制成抗体，则可用该抗体来检测以融合蛋白方式表达该蛋白的克隆。能表达出有活性的蛋白的 cDNA 文库还可对其生物活性进行筛选。不表达编码蛋白的 cDNA 文库可以通过翻译结合于（杂交于）克隆群的 mRNA 来进行筛选，那些能翻译出目的蛋白的克隆将再被细分。这些途径中一些将在本主题中被描述。

菌落及噬菌斑杂交

虽然 λ 基因文库产生噬菌斑而非菌落，除初始步骤外，这二种筛选方法基本相同。第一步如图 I3.1 所示是将噬菌斑或菌落中的部分 DNA 转移到一张**尼龙膜**或**硝酸纤维素膜**上。因为噬菌斑是已被细菌裂解的区域，所以只要将膜置于平皿上，噬菌体 DNA 就可直接与膜结合。细菌菌落必须先被裂解以释放出 DNA，通常可在置于平板表面的膜上直接培养复制物（**影印平板培养**）。膜表面的细菌会在十二烷基硫酸钠和蛋白酶液浸膜后而裂解，而原始平板将被保留用以分离相关的克隆（即含有重组质粒的菌落）。λ 噬菌体可从噬菌斑平皿的剩余的噬菌斑中获得分离。两种情况下膜上的 DNA 都是被碱变性成单链，再通过烤膜或紫外光照射单链将与膜牢固结合。然后膜被浸于含有通常为放射性标记（参见 J1）的核酸探针的溶液中，保温以使探针与其互补序列进行杂交。杂交后，充分冲洗以除去膜上未杂交的探针，显影后有探针的区域就可显示出来。如果探针为放射性标记的，则可由对 X 线片曝光（**放射性自显影**）来进行；如果探针是用修饰核苷酸标记的，则可用抗体或酶及底物溶液来检测。通过将膜与原始平板进行比较，标出被杂交的区域，可以鉴别出最初的菌落或噬菌斑。这些菌落或噬菌斑再以较低的密度重新涂板，重复杂交过程直至获得单克隆。

表达筛选

将 cDNA 克隆至 λ gt11 载体的 *Eco*R Ⅰ 位点，该 cDNA 有 1/6 的概率能以正确的方向、正确的可读框架插入（参见 P1），翻译出它的基因产物。*Eco*R Ⅰ 位点靠近 *lacZ* 基因的 C 端，插入片段的编码区必须以正确的方向和可读框架与 *lacZ* 相连才能表达出含 cDNA 基因产物的融合蛋白。β-半乳糖苷酶融合蛋白可能含有能被天然蛋白的**抗体**所识别的多肽区域（**抗原决定部位**）。这些抗体可以用于筛选**表达文库**。筛选流程与噬菌斑杂交程序相类似，因为“噬菌斑复制”是通过在噬菌斑平皿表面放置一张膜来进行的，只不过在抗体筛选中所要检测的是 cDNA 编码的蛋白质而不是 DNA 本身。膜被处理后与蛋白质共价结合，再浸入含抗体的溶液中。当抗体与其抗原决定部位结合后，可以由其他抗体和（或）可识别它的化学物质来检测。通过这种筛选可以将表达噬菌斑的位置确定下来。需要进行重复循环筛选以获得纯噬菌斑。

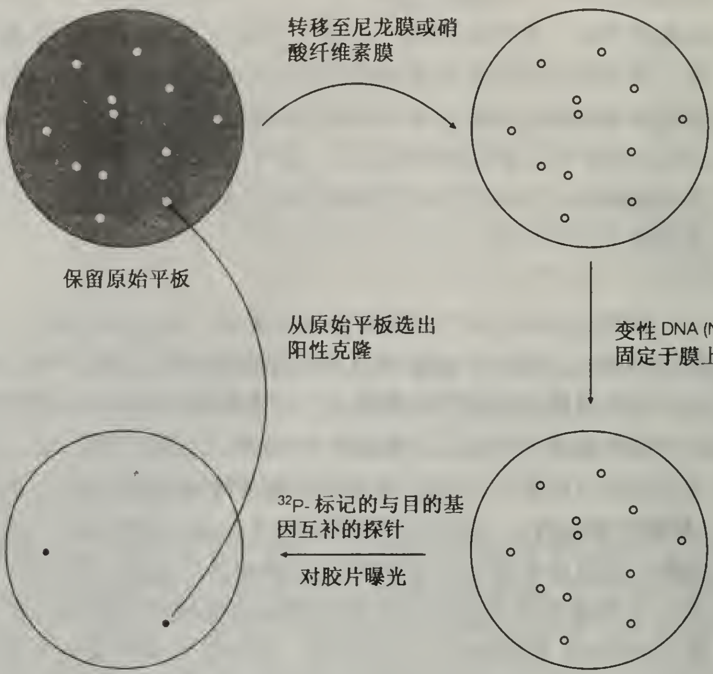


图 I3.1 经噬菌斑杂交的筛选

杂交扣留与释放

单个 cDNA 克隆或克隆群都可用来与 mRNA 样品进行杂交。杂交后 mRNA 可被直接翻译，通过检测其产物会显示出某些产物的翻译被抑制了（**杂交扣留翻译**）。将 cDNA 群细分成少量的 cDNA 组并重复杂交实验最终可以检测出抑制某蛋白翻译的单个 cDNA。另一方面杂交后，也可以纯化出杂合子（例如若 cDNA 与磁珠结合或是与引导 cDNA 合成的修饰核苷酸的抗体凝集），再将杂交的 mRNA 从杂合子释放出来（通过加热和（或）变性）并进行翻译。这种**杂交释放翻译**可鉴定出 cDNA 克隆编码的蛋白。

染色体步移

从文库中找到相邻的基因组克隆常常是必要的，也许是因为克隆的仅是基因的一部分，或是含有调控序列的 5′端或 3′端的丢失。有时可在制作遗传学图谱的实验中发现某一特定基因邻近一个已被克隆的基因，因此很有可能通过不断从文库中分离相邻的基因组克隆来克隆目的基因（**位置克隆**）。这被称为**染色体步移**，通过染色体步移可以获得代表某一特定染色体的较长连续区段的重叠基因组克隆群（图 I3.2）。要分离某一重叠基因组克隆，需要从插入片段末端制备探针。这可能会涉及该克隆的**限制性酶切图谱**（参见 J1），以便能从凝胶电泳回收某个特定片段，但某些载体可通过对载体-插入片段连接处的体外**转录**合成探针。这样的探针可以被用于菌落或噬菌斑杂交对文库进行再筛选，然后可经分析并定出新提取的

克隆与初始克隆间的相对位置。如果幸运，在几个新的克隆中，会有某个克隆能少量地与初始克隆重叠（例如重叠 25kb 中的 5kb），然后又可以用第二个克隆的远端（离初始插入片段末端最远的）制备的探针重复整个流程，并以此往复。

如果染色体步移终止，可能是因文库中不包含一段适合序列，用**染色体跳查法**可克服这一问题。当“步移”距离很大时，也能应用染色体跳查，例如当从一个相对距离起始位点进行位置克隆时。一个长 DNA 片段的相对的末端在该 DNA 环化或克隆进入一种黏粒载体（见 H3）时连接到一起。不用图谱完整的 DNA，末端片段能够用作（亚克隆）连续的探针来筛选一个文库且它能允许 50kb 左右的克隆互相之间分离开，因此形成一种比图 I3.2 所示方法更快迅的“步移”。

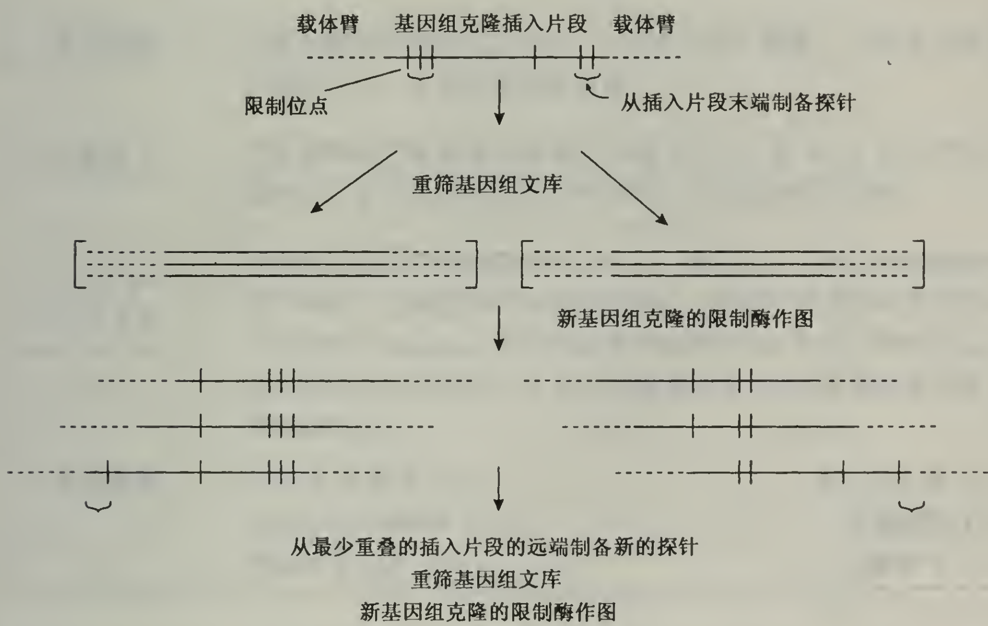


图 I3.2 染色体步移

（叶 彬 译 李文君 校）

J1 克隆的鉴定

要 点

鉴定	克隆的鉴定过程包括测定重组 DNA 分子的各种特性如大小、限制图谱、基因的方向以及核苷酸序列，该过程需要纯的已克隆化的 DNA 样品。
限制图谱	用一种或二种以上的限制酶对重组 DNA 分子进行酶解，可以构建标明该 DNA 分子的酶切位点和片段大小的限制图谱。
部分酶解	用限制酶对末端标记的 DNA 片段进行部分酶解，并测定所生成片段的大小，也可构建限制图谱。
核酸标记	可用多聚核苷酸激酶或末端转移酶对 DNA 和 RNA 分子进行末端标记。全长标记须用聚合酶合成一条完整的被标记链。
DNA 和 RNA 杂交	将凝胶泳道中的核酸带转移至膜上，固定之后与标记的核酸探针进行杂交，经洗脱除去未杂交的探针，然后对膜进行处理以显示出杂交带。Northern 杂交可检测特定的 RNA 分子，而经 Southern 杂交所显示出的 DNA 带应该是基因组 DNA 中基因或已克隆基因的部分。

相关主题

DNA 克隆概述（G1）　基因组文库（I1）
质粒 DNA 的制备（G2）　筛选流程（I3）
限制酶与电泳（G3）　核酸测序（J2）

鉴定

鉴定基因组克隆或 cDNA 克隆的过程应从纯的 DNA 样品的制备开始，然后才是鉴定该克隆的部分或全部的特性，包括片段的大小、插入片段的大小以及插入 DNA 片段的限制酶切图谱（**限制图谱**）。其中是否含有基因（转录序列），如果有则该基因的所在位置及其方向以及插入 DNA 的部分或全部序列。在未能快速进行 DNA 测序（参见 J2）之前，鉴定的步骤通常大致按上述顺序进行，但现在对克隆进行部分或全部测序往往是第一步。虽然序列分析可提供推测基因的大小、限制图谱以及方向或者**可读框**（**ORF**）（参见 P1），但仍须有实验证明推测基因确实在某生物体的一些细胞中转录成 RNA，而对于那些编码蛋白的基因还要证明所预测大小的蛋白确实在体内被合成。从细菌菌落进行质粒 DNA 的制备已在 G2 中叙述。除此之外 DNA 也可很方便地从噬菌体颗粒的载体中制得。这一过程包括：先用**噬菌斑经纯化的噬菌体**在允许细胞裂解的条件下感染细

菌，然后从裂解液中移去细胞残骸，经直接离心或在高盐溶液中的聚乙二醇沉淀来从细胞裂解液中回收噬菌体颗粒。再经苯酚-氯仿抽提、乙醇沉淀，最终得到的 DNA 通常溶解在 TE 缓冲液中（10mmol/L Tris·HCl，1mmol/L EDTA，pH 8.0）（参见 G2）。

限制图谱

通过琼脂糖凝胶电泳并与已知片段大小的 DNA 标记相比较可以测定线状 DNA 分子的大小（参见 G3）。对质粒或噬菌体克隆中插入的基因组 DNA 片段或 cDNA 片段的大小，最好用能将插入片段与载体 DNA 分开的限制酶来酶解。对用 *Eco*RⅠ位点构建成的 cDNA 克隆，还用 *Eco*RⅠ来酶解将可回收到插入片段与载体。将酶切后的样品与 DNA 标记一起进行琼脂糖凝胶电泳就可测得插入片段及已知载体 DNA 的大小。这一信息只是制作重组 DNA 分子图谱所需信息的一部分，而载体中**插入片段的方向**（参见 H2），以及若是多个片段则其排列顺序都是未知的。要获得这些信息，可以用不同限制酶来单切、或组合起来进行切割。图 J1.1a 给出了一个 λ EMBL3 基因组克隆限制图谱的制作过程。*Sal* Ⅰ 在 *Eco*R Ⅰ 或 *Bam*HⅠ克隆位点附近切割并从载体上释放出插入片段。因此在插入片段内没有 *Sal* Ⅰ切割位点，所以仅形成大小约为 9kb、15kb 和 19kb 的三个片段。在载体短臂上只有一个 *Hin*dⅢ位点，在距末端约 4kb 处。在这一例子中，*Hin*dⅢ在插入片段中还有两个切点，共产生大小分别为 4kb、7kb、11kb 和 21kb 的四个片段。4kb 是短臂的一部分，21kb 片段必定是长臂加上 2kb（19 + 2 = 21），但剩下两个片段的顺序仍然不能确定。用 *Sal* Ⅰ与 *Hin*dⅢ的**双酶切**产生出 2kb、4kb、5kb、6kb、7kb 和 19kb 六个片段，其中 19kb 以及 4kb 和 5kb 一起是载体的两个臂。由于 *Sal* Ⅰ在克隆位点处切割，所以会将 21kb 切成（19 + 2）kb，11kb 切成（6 + 5）kb，其中 6kb 是插入片段的一部分。7kb *Hin*dⅢ片段未被 *Sal* Ⅰ切割，表明该片段位于插入片段的中间，结果构成了如图 J1.1a 所示的限制图谱。

部分酶解

当某种限制酶的切点较多，且双酶解产生的图谱又很复杂时，**部分酶解**技术可提供制作完整限制图谱所需的信息。这一过程最好是在 DNA 片段的一个末端加上一个修饰过的、或具放射活性的核苷酸，形成一个如图 J1.1b 所示的**末端标记分子**。用 *Eco*RⅠ完全消化未标记的 10kb 片段产生出 1kb、2kb、3kb 和 4kb 四个片段，而部分酶解能产生出所有可连续的片段，当总 DNA 被酶解后再经 EB 染色显示出 6kb、7kb 和 10kb 的三条增加条带。注意到 3kb、4kb 和 6kb 的片段可由多种方式产生，但部分酶解泳道被放射自显影来显示那些末端标记的分子，切割图谱应该是 3kb、4kb、6kb 和 10kb。这些信息就足以构建出如图所示的图谱。

核酸标记

可以对 DNA 和 RNA 分子的末端进行标记（**末端标记**）或全长进行标记（**均匀标记**）。从一端开始均匀标记也是可能的，但在有限的时间内，

并非产生一个全长均匀标记的分子，其方式可能更像末端标记分子。也可以只标记双链分子的 DNA **特定链**。标记通常用放射性同位素进行，但也有非放射性标记方法，例如可被生物素特异抗体检测的生物素标记 dUTP 来标记。

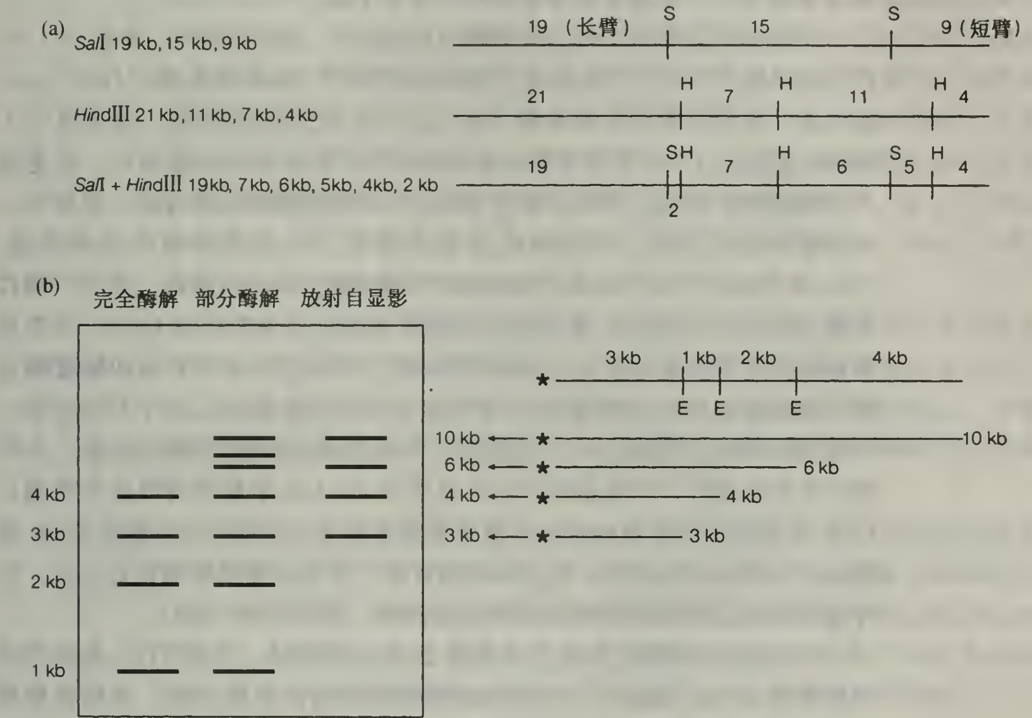


图 J1.1 限制图谱。(a) 单酶解和双酶解完成的图谱；(b) 经末端标记分子的部分酶解所制作的图谱。* 代表标记末端。

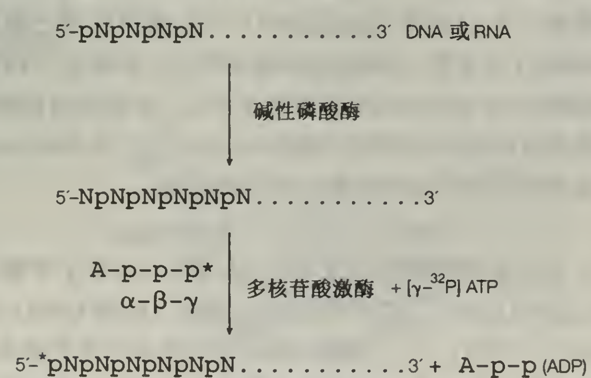


图 J1.2 核酸分子的 5′端标记

5′端标记（图 J1.2）可用**多聚核苷酸激酶**（**polynucleotide kinase**）来

进行，结果是将具放射活性的磷酸加到具有游离 5′-OH 的核酸上。游离 5′羟基可由**碱性磷酸酶**（alkaline phosphatase）去**磷酸化**（dephosphorylation）产生。ATP 的最外端的 γ-磷酸是用来标记的标记源（参见 G1 中的表 1）。对于 **3′端标记**则可用**末端转移酶**（terminal transferase）向核酸的 3′端加上一个或多个核苷酸进行（参见 I2）。

诸如由某些限制性内切核酸酶（*Eco*RⅠ、*Bam*HⅠ等，参见 G1）所产生的双链 DNA 分子的凹进 3′端可通过多种 **DNA 聚合酶**（DNA polymerases）采用被标记的核苷酸（如［α-^{32}P］dCTP）来填外。因为每一 3′端只能被加上 1～4 个核苷酸，该 DNA 分子基本上是末端标记，但是是不同链的两个末端。若对这样的 DNA 分子进行限制图谱构建，须用另一种限制酶进行酶切，再分离出各具末端标记的片段来制作限制图谱。DNA 聚合酶也可被用来制作均匀标记的**高活性**的 DNA 探针。在进行**切口平移**（nick translation）标记中，用微量 DNase Ⅰ处理双链 DNA，会在其两条链上引入随机的切口。DNA 聚合酶Ⅰ利用它的 **5′→3′外切核酸酶**活性可找到这些切口（参见 E2）并移去 dNTP，随着从切口的 5′端移去一个残基，将在 3′端加上一个核苷酸，从而利用它的聚合酶活性掺入具放射活性的核苷酸，结果是切口的位置沿该 DNA 分子移动（即定向平移）。DNA 聚合酶也可在所有组合六聚核苷酸均存在下，在变性双螺旋 DNA 模板后，被用来制备探针。这些六聚核苷酸于单链上的互补位置处退火，并作为聚合酶合成互补放射活性的链的**引物**。亦见 PCR（J3）。

可以通过将双链 DNA 分子克隆于 M13 噬菌体（参见 H2）载体而获得的单链 DNA，或经 **3′→5′外切核酸酶**酶解后的单链片段，来制备链特异 DNA 探针。DNA 聚合酶可以像合成第二链 cDNA 那样再合成另一掺入具放射活性的核苷酸的链（参见 I2，图 1a）。链特异 RNA 探针可由 RNA 聚合酶如 **SP6、T7 或 T3 噬菌体聚合酶**（参见 H1）**体外转录**来制备。如果将目的序列克隆进一个在克隆位点的末端有上述任一聚合酶启动子的体外转录载体（如 pGEM、pBluescript）上，就可以用一种聚合酶制得**有义 RNA** 转录物（与天然、RNA 转录物相同），而用另一种聚合酶制得**反义 RNA** 转录物（与天然 RNA 转录物互补）。具放射活性的 NTP 被用于标记。链特异探针的反义链可与细胞 RNA 杂交，在 Northern 杂交（参见下文）中很有用，因而有义链探针将用作对照。

DNA 与 RNA 杂交

为了检测琼脂糖凝胶上众多 DNA 或 RNA 分子中能与特定探针杂交的 DNA 或 RNA 分子，进行 **Southern 杂交**（检测 DNA）或 Northern 杂交（检测 RNA）（图 J1.3）。前者根据其发明者的名字来命名的，而后者是由前者推导而来。核酸子经**琼脂糖凝胶**电泳分离，再转移到尼龙膜或硝酸纤维素膜上。这一步通常由毛细管作用实现的，也可由电转移、真空转移或离心来完成。就 Southern 印迹而言，转移前 **DNA** 通常用**碱变性**（参见 C2），形成单链后方可用于杂交。而对于 RNA，并不需要进行碱变性，碱

变性反而会水解 RNA 分子（参见 C2）。转移后 DNA 杂交和 RNA 杂交的步骤就一样了。核酸被固定在膜上，与标记探针杂交，充分洗涤后，检测出杂交的探针（通常用放射自显影或抗体法）。杂交和洗脱的条件很关键，如果探针与目的序列 100% 相同，应进行高严紧杂交。严紧性（stringency）是由杂交缓冲液中的盐浓度和杂交温度来决定的（高温和低盐的条件是最严紧的，只有完全匹配的杂交分子才稳定）。而对与目的序列不完全匹配的探针，严紧性应降到可形成不完全匹配的杂交分子的程度。如果杂交或洗脱的严紧性太低，探针就会与许多序列结合，而降低特异性。向杂交缓冲液添加甲酰胺可将实际杂交温度降低约 25℃，从通常的 68℃ 降至 43℃。洗脱步骤应在比探针和目的序列的理论融化温度 T_m 低 12℃ 的条件下进行，T_m 可由公式 $T_m = 69.3° + 0.41\ [\ (G+C)\%\ ]\ - 650/l$ 求得，其中的 l 表示探针分子的长度。

RNA 杂交能提供有关 mRNA 及任一前体大小的信息，还可用来判定用作探针的 cDNA 克隆是否是全长，或者是否是相关转录产物家族中的一种。RNA 杂交还可有助于确定一基因组 DNA 克隆是否含有转录区，当被转移的 RNA 是来自不同组织时，还可确定这些转移物是在哪种组织中被转录。

基因组 DNA 克隆片段的 Southern 杂交可用 cDNA 分子为探针来找出基因组克隆的哪一部分与 cDNA 片段相对应。如果用包含整个基因组的 DNA 片段来进行 Southern 杂交，结果会给出有关基因组中特定基因片段的大小，以及该基因在基因组中有多少拷贝的信息。来自不同生物体的 DNA 或 RNA 样品的杂交会显示出一个基因在不同种间的保守程度。

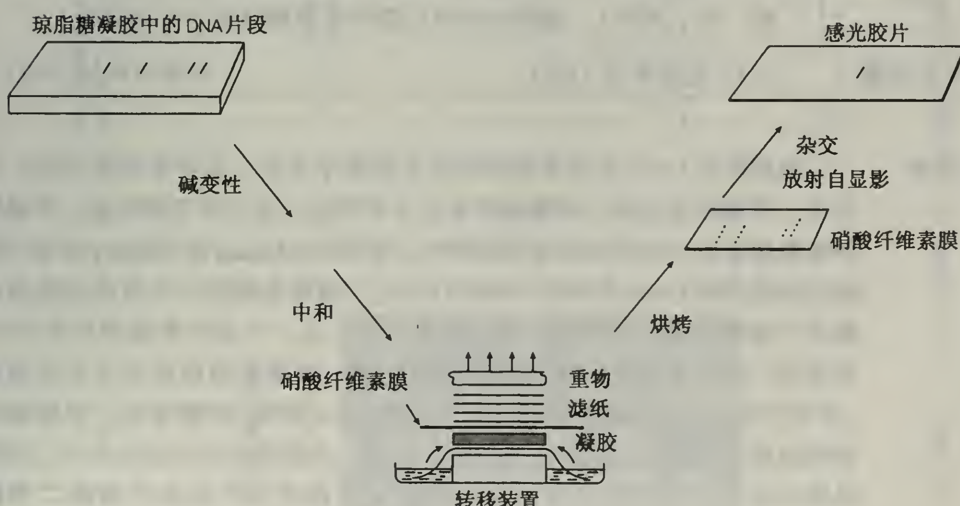


图 J1.3 DNA 杂交

J2 核酸测序

要 点

DNA 测序	DNA 测序有两种主要方法，一是 Maxam 和 Gilbert 的化学法，即在对末端标记的 DNA 进行碱基特异切割反应后进行凝胶分离的方法；另一是 Sanger 的酶学法，即用双脱氧核苷酸作为链终止试剂，通过聚合酶的引物延伸产生一系列大小不同的分子后再进行分离的方法。
RNA 测序	可在特定核苷酸的 3′端切割的一套四种 RNase 被用来消化末端标记的 RNA 形成一系列大小不同的片段，再经聚丙烯酰胺凝胶电泳最终读出序列。
序列数据库	新测定的 DNA、RNA 以及蛋白质序列被输入数据库（EMBL 和 GenBank），所有进入数据库的已知序列都可随时被调出用于计算机分析。
序列分析	用特定的计算机软件来对核酸及蛋白质的序列进行检索与分析，以找出某种图谱（如限制酶切位点）或相似性（如对新序列）。
基因组测序计划	几种生物的全部基因组序列已被测定（病毒、细菌、酵母、线虫、拟南芥和人），而另一些生物的基因组测序正在进行中（水稻、鱼、鼠等）。通常先制作出遗传学图谱以助于计划进行。
相关主题	DNA 克隆概述（G1）　　噬菌体载体（H2）

DNA 测序　测定较长 DNA 分子序列的两种主要测序方法（化学法和酶学法）的任何一种都涉及具有一末端相同而大小不同的一系列分子的形成，再经聚丙烯酰胺凝胶电泳分离以读出其序列。早期的 **Maxam 和 Gilbert** 的化学法要求待测序的 DNA 片段的一端进行标记，通常是添加一个放射活性的磷酸至 5′端或 3′端，或加上一核苷酸至 3′端。这一方法对单链和双链 DNA 都适用，并涉及两步反应的**碱基特异切割**。碱基先用特定化学试剂修饰（参见下文），然后**哌啶**可在此位点切割 DNA 的糖-磷酸骨架。控制碱基修饰的反应时间或反应物浓度可确保生成一个系列依次加长的分子，而不是完全切成短的寡聚核苷酸。特定的碱基修饰反应包括用**硫酸二甲酯**(dimetlyl sulfate，DMS) 甲基化 G 碱基，**甲酸**作用于嘌呤 A 和 G，而**肼**(hydrairne) 可被用来在嘧啶处（C+T）进行水解，但高盐浓度会抑制 T 反应。所以对测序凝胶的四个泳道（G、A+G、C+T 和 C）进行分析就可以确定出序列（图 J2.1a）。 这一方法适用于对未克隆过的基因组 DNA

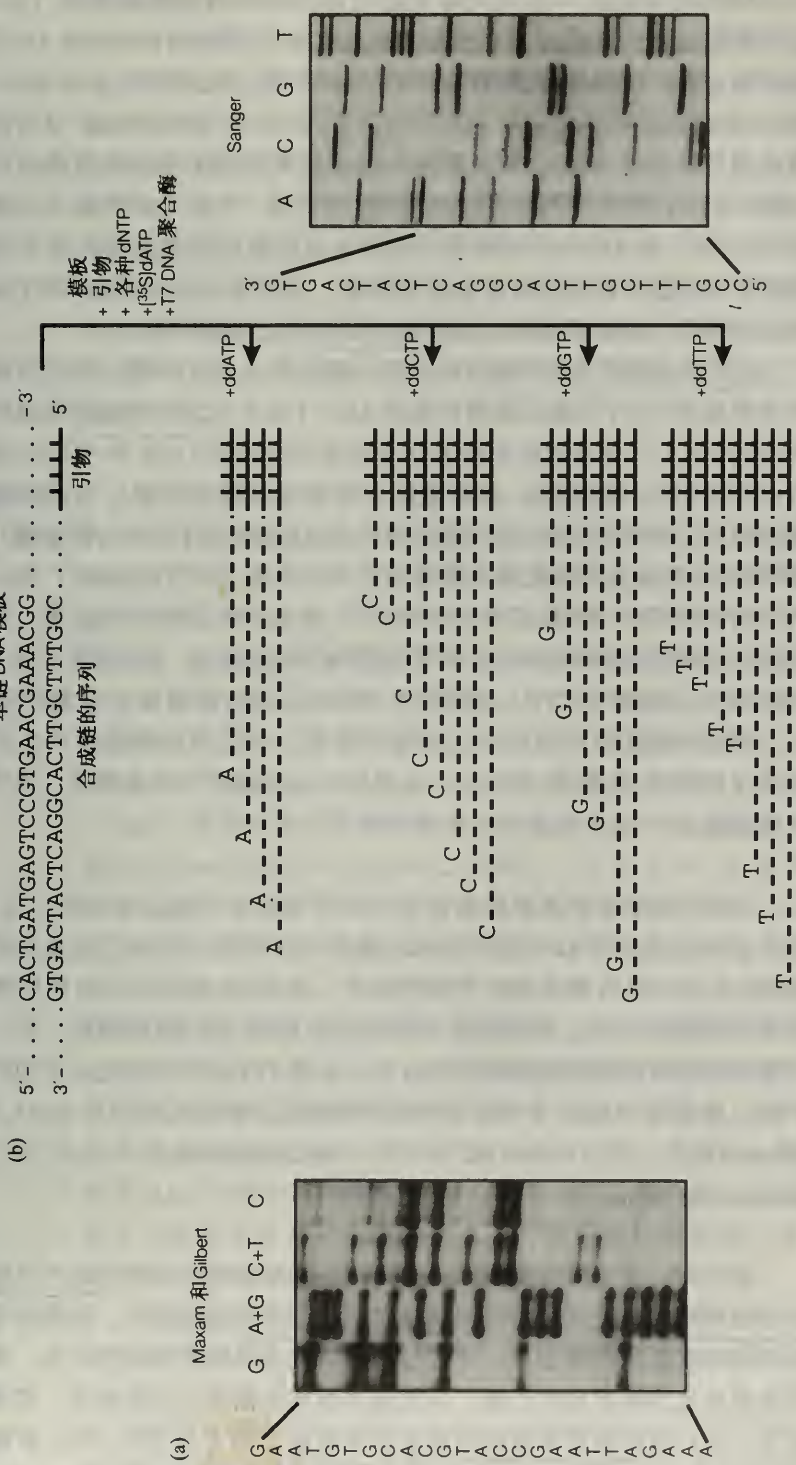


图 J2.1 DNA 测序。(a)Maxam和Gilbert序列胶的一个例子；(b) Sanger 测序法图解

测序。

DNA 测序的化学法很大程度上已被用四种**双脱氧核苷酸**（ddNTP）进行末端终止的 **Sanger** 方法代替（图 J2.1b）。测序引物与单链 DNA 模板分子结合后，DNA 聚合酶用 ddNTP 延伸引物。延伸反应分成四组进行，每一组分别用四种 ddNTP 中的一种来进行终止，再用 PAGE 分析四组样品（通常具放射活性）。双脱氧核苷酸在脱氧核糖上没有聚合酶延伸链所需的 3′-OH 基团，所以可被用作链终止试剂。可在合成步骤通过掺入诸如［α-^{35}S］dATP 来进行标记，或者也可先将引物末端用具放射活性的或荧光染料进行标记后再进入合成步骤。后者常在自动 DNA 测序仪进行测序时使用，最常使用的是具荧光标记的双脱氧核苷酸。

这一方法需要使用单链 DNA 作为模板来合成互补链，这就意味着在测序前要将 DNA 克隆进噬菌体载体 M13（参见 H2）。从噬菌体回收得到的单链 DNA 可与在载体和插入片段连接区附近的 DNA 序列互补的 15～17nt 长的引物一起**退火**。所有克隆进该载体中的序列都可用这一通用引物来测序。接着将 DNA 聚合酶（通常是 **Klenow** 或 **T7 DNA 聚合酶**）及少量的其 α-磷酸上一个氧原子被硫原子所代替的［α-^{35}S］dATP 加入退火的引物和模板中（如果引物未被标记）。然后分注进四个反应管，每个均含有特定比例的某种链终止 ddNTP 及通常的 dNTP（如试管 C 含有 ddCTP 以及 dATP、dCTP、dGTP 和 dTTP）以保证只有一定量的链被终止。将这四组反应产物用 PAGE 进行分析，与化学法相比，产生的假带较少（比较图 J2.1a 与 J2.1b）。双脱氧方法已进行了许多改进，现在可用双链模板或 PCR 产物进行（参见 J3）。

RNA 测序

尽管 DNA 由于其较好的稳定性及可靠的基于酶的分析法，比 RNA 测序容易得多，但有时直接对 RNA 进行测序还是必要的，尤其是确定诸如 tRNA 或 rRNA 被修饰核苷酸的位置（参见 O1 和 O2）。用能切割 3′端特异核苷酸的 RNase 来酶解具 5′端标记的 RNA 进行序列测定。这一反应也需要控制限制酶的用量和反应时间，以期产生出一系列大小不同的切割产物，然后用 PAGE 分析。下列四种 RNase 常被用来进行 RNA 测序，**RNase T1** 在 G 后、**RNase U2** 在 A 后、**RNase Phy M** 在 A 与 U 后、***Baccillus cereus*** **RNase** 在 U 和 C 后进行切割。

序列数据库

多年来，在世界范围内已经测定出大量核酸序列，并且绝大多数科学期刊都要求在接受发表前先将其核酸序列送至公共数据库。数据库管理者将与研究者共享这些信息，并让公众可以很容易地获取这些信息，使得这些数据库成为极有用的资源。目前新序列在数据库中不断增加，越来越需要专一的计算机软件来更好地利用这些数据。两个最大的 DNA 数据库是欧洲的 **EMBL** 和美国的 **GenBank**。此外也建成了蛋白质及 RNA 序列的数据库，而一些公司还有他们自己的私有序列数据库。

序列分析

当 cDNA 或基因组克隆的序列被测定后，不经对该序列的检索与分析很难立刻找到该序列的特征。对于 cDNA 克隆，如果该 cDNA 是用多聚 dT 为引物构建成的，则该序列的一头应含有一串 A 残基。如果一串 A 存在，就可以根据多聚 A 应在 3′端来确定克隆的方向。但其他特征就难以直接用肉眼确定，并且基因组克隆并不含可用来鉴定方向的多聚 A 序列。所得序列通常用计算机和软件包来进行分析，例如**威斯康星大学的 GCG 软件包**，或 PC 程序如 **DNAstar**。这些程序有两个主要用途：一个是可鉴别出重要的序列特征，如限制酶切位点、可读框、起始与终止密码子以及潜在的启动子部位、内含子-外显子连接区等；而另一个则是将新序列与数据库内所有已知序列进行比较，以期判定是否曾获得过相关序列。**生物信息学**（bioinformatics）就是用来描述用于分析生物数据的软件开发与应用的术语。

基因组测序计划

替代标准的 ddNTP，**自动 DNA 测序**使用 4 种不同的、荧光染料标记的链终止物。该方法与传统测序相比获得了更大收益，因为激光读取不同颜色的序列在真实的时间，直接从胶的底物开始，而不需要放射自显影的时间消耗。而且，所有的 4 种反应可以在同一试管内完成并上样在同一条泳道内，且**机械手工作站**能同时准备和处理多个样品。这些发展可允许许多生命体的完整基因组序列得到测定。已完成测定的基因组包括噬菌体和病毒如 λ 及 AIDS 病毒（HIV）；超过 20 种细菌包括 *E. coli*（4.639×10^6bp）及 *Helicobacter pylori*，一种引起溃疡和胃癌的致病菌（1.57×10^6bp）；真核生物包括酿酒酵母 *Saccharomyces cerevisae*（12.1×10^6bp），线虫 *Caenorhabditis elegans*（100×10^6bp）以及果蝇 *Drosophila melanogaster*（160×10^6bp）。许多其他基因组测序计划正在进行中，包括植物、鱼、鼠、人。在最近的报道中，22 号染体（34.5×10^6bp）到 1999 年底已有 97%完成，全基因组约 90%的草图将有望在 2000 年中期完成。所谓“**完成**”问题是填补很难被克隆和（或）测序的缺口（如人 22 号染色体序列上的 11 个小缺口）。完成及全部破译人类基因组序列可能还需几年时间，但先前已构建好的详细遗传学图和一套覆盖 23 条染色体的每一条的重叠基因组克隆群将有助于计划最终完成。大部分人类基因组的测序使用 BAC 克隆而不是 YAC 克隆（参见 H3），是因为已证实后者在增殖过程中不稳定且含有不连续的插入。一种可选择的策略：随机（shotgun）测序策略依赖于巨大的计算能力来组装随机产生的序列。[1]

完成全基因组序列的可能性大大促进了**基因组学**（研究生命体基因组的学科）的发展，因为现在构成一个生命体所需的全部基因的数量、位

1）拟南芥的全基因组序列已于 2000 年完成，人类基因组 97%的序列已于 2001 年初完成，可望在 2001 年底完成全序列测定。随着技术的发展，策略的成熟，基因组计划的进程已大大加快。——译者注

置、总体大小以及组构都能知道。而且分布于全基因组上的多种多样的卫星序列（参见 D4）以及拟基因的位置也将帮助我们理解基因组进化。或许更重要的是能够通过比较不同生命体基因组所获得的信息。例如在 *S. cerevisae* 中，从基因组序列预测至少 6200 个基因，其中约一半是未知功能。当前所面临的挑战是努力发现通过基因组测序计划预测的大量未知基因的功能。这就需要大规模基因失活方法，也就是**功能基因组学**，以及与之相伴的蛋白质组学方法（参见 B3）。

在自动 DNA 测序中更进一步的进展是可较好地利用 **DNA 芯片**技术，该技术是在一种固相支持物如玻璃或尼龙上高密度排列不同的 DNA 序列。像较低密度排列的 **DNA 微阵列**一样，它们均被用于平行杂交实验中来检测哪种序列存在于某一混合物中。应用于 DNA 测序，DNA 芯片须包含一个特定长度的每种可能的寡核苷酸序列，因为最大序列“读取”可能性是芯片上寡聚核苷酸数量的平方根。因此全部 65536 个 8-mers 可允许最大读出序列为 256bp，而超过 10^{12}个 20-mers 将可以读出 1Mb 序列。因此，每平方厘米排列约 10^6 寡聚核苷酸现在是可能的，在读取超过现在 1kb 限制之前还需技术上的明显改进。

J3 聚合酶链反应

要 点

PCR	聚合酶链反应（PCR）利用与 DNA 模板序列的两端互补的一对寡聚核苷酸引物来扩增一段 DNA 序列。由一种热稳定的 DNA 聚合酶经三步反应即变性、引物退火和聚合的循环从两个引物来相对延伸。
PCR 循环	反应循环由三步组成：95℃变性双螺旋 DNA，约 55℃退火使引物与模板结合，以及 72℃聚合反应。除模板、引物、缓冲液和酶外，还需要 Mg^{2+} 和 dNTP。
模　板	几乎所有含有一个或多个完整的目的 DNA 分子的 DNA 在理论上均可用作 PCR 扩增的模板，其条件是只要能设计出合适的引物。
引　物	(G+C) 含量相近的 18～30nt 的一对寡聚核苷酸，只要能引导 DNA 的相向合成，可作为 PCR 引物。如果目的 DNA 序列不完全清楚，也可以使用具简并性的引物。
酶	热稳定的 DNA 聚合酶（如 *Taq* 聚合酶），因能经受热变性步骤而应用于 PCR，但其中一些较易出错。
PCR 的条件优化	须通过改变退火温度或 Mg^{2+} 浓度来获得真实的扩增结果。对于复杂的混和物，采用另一对嵌套引物可改善其特异性。
PCR 派生技术	在基本 PCR 基础上的派生技术包括定量 PCR、简单寡核苷酸引物 PCR（DOP-PCR）、反向 PCR、复合 PCR（multiplex PCR）、cDNA 末端快速扩增（RACE）及 PCR 诱变。
相关主题	DNA 克隆概述（G1）　克隆的鉴定（J1） 基因组文库（I1）　克隆技术的应用（J6） 筛选流程（I3）

PCR　如果能设计一对**寡聚核苷酸**引物与目的 DNA 分子互补，以致于能被 DNA 聚合酶相向延伸，那么被该引物结合的模板区就可通过**变性**、**引物退火**和**聚合**的循环来大量扩增。这一过程被称为**聚合酶链反应**（polymerase chain reaction），已成为分子生物学中一种在克隆及基因分析中的必需工具。**热稳定的 DNA 聚合酶**的发现使得 PCR 循环的步骤更为方便，

目前已在许多科学领域得到广泛应用（参见 J6）。

PCR 循环

PCR 的工作程序如图 J3.1 所示：在第一个循环中，目的 DNA 在加热至 **95℃**，典型的为 60s 左右的条件下分成两条链，当降温至 **55℃**（约 30s）左右时，引物实现与模板 DNA 退火。实际退火温度依引物长度和序列而定。退火后再升温至 **72℃**（60～90s）以进行最后聚合反应。这一步要消耗反应混合物中的 **dNTP**，并需要 Mg^{2+}。在第一次聚合反应中，不同的目标分子从引物位点开始扩增不同长度，不断对目标分子进行拷贝，直至温度再次升到 95℃，变性新合成分子的第二次循环开始。第二次退火时，反应液中的另一引物与新合成链结合，经聚合反应拷贝模板至第一个引物的末端。所以在第二次循环后，就合成出了新的正确长度的分子。在之后的循环，这些分子数目迅速超过最初的模板分子，每次循环均以翻一番的速度增长。如果 PCR 的效率是 100%，n 次循环后单个目的分子就会扩增至 2^n。实际中通常循环 20～40 次。

模板

因为可实现大量的扩增，有时在少至一个起始模板分子的情况下也能实现 PCR 扩增。因此任何可提供一个或多个目的分子的 DNA 原则上都能用作 PCR 的模板，包括血液、精子或任何其他组织，如较早的法医标本、古生标本制备的 DNA、或者从实验室里的细菌菌落或噬菌斑制备的 DNA 以及已纯化的 DNA。无论模板 DNA 来源如何，只要已知一些序列信息从而能设计出引物的 DNA 就能应用于 PCR 扩增。

引物

每一对 PCR 引物需 18～30 核苷酸，并且在一对引物间的 **G+C 含量**相似，以致使它们在相近的温度下与其互补序列结合。对于短的寡核苷酸（<25nt），退火温度（℃）可用下式计算：$T_m=2(A+T)+4(G+C)$。引物被设计成可与目的序列的相应链退火，结果可通过向 3′端加核苷酸相向延伸。较短的目的序列较易扩增，因此常扩增小于 500bp 的片段，但在最适条件下 PCR 能扩增大于 10kb 的片段。如果待扩增的 DNA 序列是已知的，则引物设计相对容易。对于待扩增的区域，要在该区域的两边均找出约 20 核苷酸、(G+C) 含量相近的适合序列（如某些癌症中的突变位点，参见 J6）。如果 PCR 产物要用于克隆，最好在引物 5′端加上单一的限制酶切位点。

如果目的 DNA 序列是未知的，例如克隆一个只知道部分氨基酸序列的蛋白的 cDNA，设计引物就会比较困难。因此可通过遗传密码（参见 P1）解出编码已知氨基酸序列的 DNA 序列，再设计出**简并引物**。例如 His Phe Pro Phe Met Lys 由 5′-CAYTTYCCNTTYATGAAR-3′的 DNA 序列编码，（其中 Y=嘧啶、R=嘌呤、N=任意碱基。）该序列具有 $2\times2\times4\times2\times2=64$ 倍简并性。因此若合成一个所有 64 种序列的混合物来用作引物，其中之一必是正确的。另一引物可用类似方法制备。如果已知肽

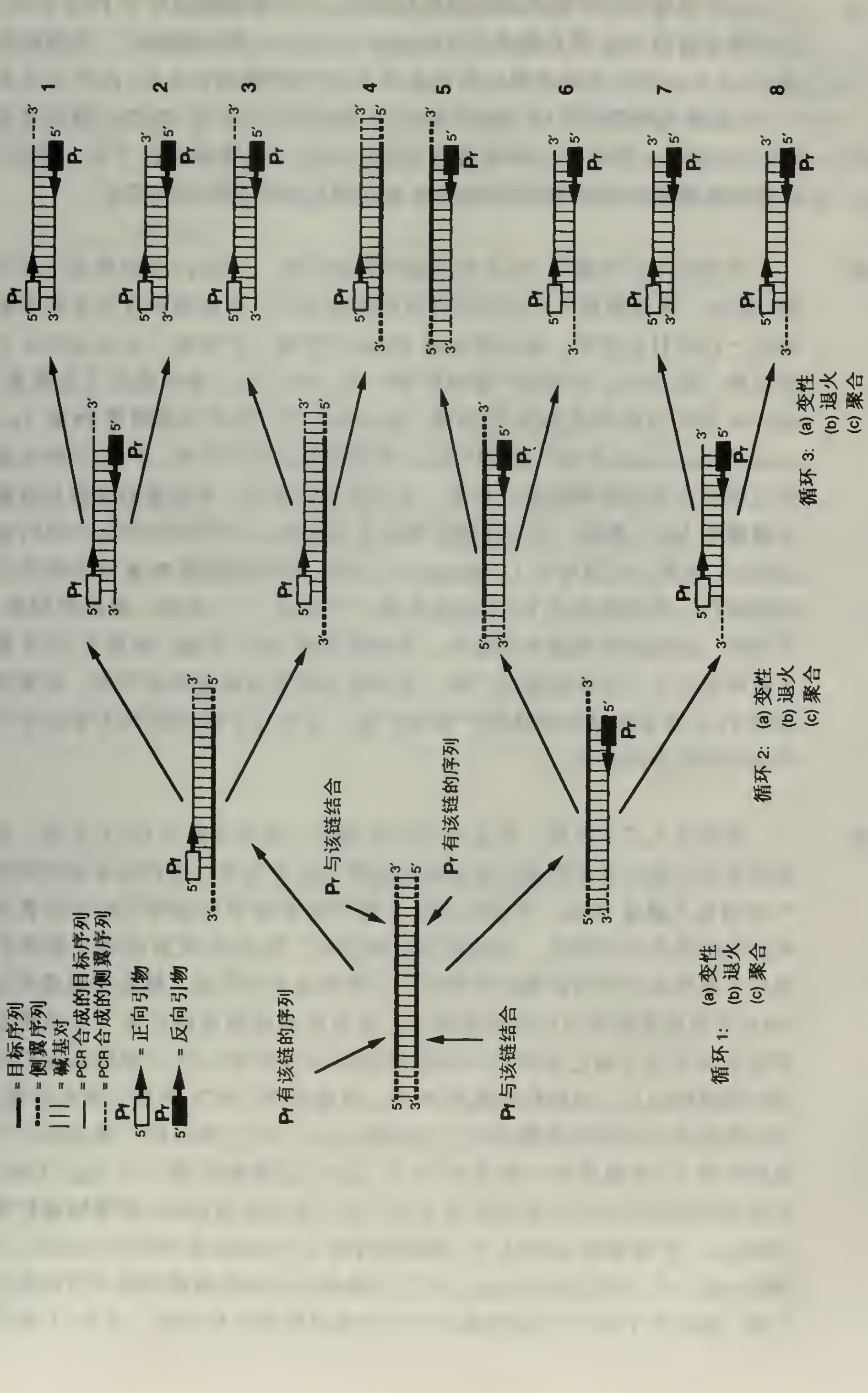


图 J3.1 聚合酶链反应的头三个循环。只有在三个循环后，才会出现扩增区域长度相同的双链分子（分子 2 和 7），再经 n 个循环后，这些就变成主要产物

链序列之一是 N 端序列，那么序列的顺序也就知道了，从而可决定引物方向。应用**简并寡核苷酸引物**的 PCR 有时亦称为 **DOP-PCR**。

酶

从许多嗜热菌中分离得到的热稳定的 DNA 聚合酶被用于 PCR 反应。其中最常用的 ***Taq* 聚合酶**是从 *Thermus aquaticus* 菌中制备的。该酶能经受 1～2 min 95℃变性步骤，在该温度下的半衰期大于 2 h。由于不具备 3′→5′的校正外切酶活性（参见表 F1.1 和表 G1.1），已知 *Taq* 酶在复制 DNA 时会引入错误，大约每聚合 250nt 发生一次错误。由于这一原因，常选用其他精度更高的热稳定 DNA 聚合酶用于某些特定的实验。

PCR 的条件优化

即使使用已克隆的 DNA 和已知序列的引物，一般 PCR 效率也不会达到 100%，通常都要优化反应条件以期提高效率。当试图从众多序列中扩增出一特定目标序列，如从基因组 DNA 中扩增一个基因，或从 cDNA 文库或第一链 cDNA 合成的产物中扩增一个 cDNA 时，条件优化尤为重要。mRNA 反转录后再用 PCR 扩增第一链 cDNA 的方法称为**反转录 PCR**（reverse transcriptase PCR，RT-PCR）。如果反应条件不适宜，PCR 产物在凝胶上就会出现不清晰的成片条带，而不是明显条带。常改变的参数包括**退火温度**和 **Mg^{2+} 浓度**。退火温度太低会发生错配。对不同序列有不同的最适 Mg^{2+} 浓度，但通常在 1～4mmol/L 之间。可用**嵌套式 PCR** 来改善反应的特异性。所谓嵌套式 PCR 就是在第二轮 PCR 中，使用一套新的引物，它与第一套引物扩增的片段结合，形成更短的 PCR 产物。如果在 PCR 第一轮中产生出一些非特异的产物，在凝胶上带不清晰或有许多带，而使用嵌套 PCR 就会确保只有目的产物被扩增，因为只有目的序列才会含有与两套引物结合的位点。

PCR 派生技术

假如加入多对引物，PCR 可用于扩增同一反应中多个 DNA 片段，且如果这些片段长度不同则可容易地在凝胶上区分开来。这种用多套引物的 PCR 被称为**复合 PCR**，并经常用于一种快速实验中以检测可能污染食物或水体的微生物的存在，或被感染组织存在。对基本 PCR 的改良使得扩增位于基本成对引物扩增的序列的上下游区域成为可能。例如如果基因组 DNA 首先用限制性内切核酸酶消化，然后用连接酶连接环化，一对背对背的引物可用于从已知序列区域扩增该圆环以获得 5′及 3′侧翼区直到连接的限制性位点。这被称作**反向 PCR**。当通过 RT-PCR 得到 cDNA 片段，可以首先利用末端转移酶加尾，例如将 oligo（dC）加到第一链 cDNA 末端来扩增 5′-侧翼序列（参见图 J3.1，I2）。这样就可用一与 oligo（dG）引物相连的基因特异引物来扩增 5′区。这一技术称为 **cDNA 末端快速扩增**（RACE）。扩增真核 mRNA 3′-侧翼序列的 3′- RACE 则利用与 mRNA 3′端的 poly（A）尾退火的 oligo（dT）引物和与之相连的基因特异引物进行扩增。通过在 PCR 反应的后期加入放射性或修饰的核苷酸，或如 J1 描述

的那样标记 PCR 产物，PCR 可以用于制备标记的探针筛选文库（参见 I3）或进行杂交试验。PCR 还可用于将特异突变引入到一特定 DNA 片段中，J5 中介绍了一个关于 **PCR 诱变**过程的例子。

定量 PCR 可以确定一实验样品中的 DNA 数量（分子数）。定量 PCR 最好方法之一是在扩增前的样品中加入一已知数量的相似 DNA 片段，如含有短的缺失的 DNA 片段。两种产物的比例取决于加入的缺失片段的数量，并可算出样品中目标分子的数量。在**不对称 PCR** 中，只有一条链被扩增（以线性方式），当应用到 DNA 测序时（参见 J2）被称为**环状测序**(cycle sequencing)。PCR 还可用于增加 DNA 指纹分析的灵敏度（参见 D4 和 J6）。

J4 克隆基因的组构

要 点

组 构	以寡脱氧胸腺苷酸引导的 cDNA 克隆的极性可以很容易地从 poly (A)的位置来判断，由此推导出编码区。基因组克隆中基因的存在和其极性不很明显，但可通过作图和杂交实验来确定。
在基因组 DNA 上定位 cDNA	以 cDNA 克隆的一部分为探针进行 Southern 杂交，可显示出基因组克隆中的哪一部分对应于 cDNA 序列。
S1 核酸酶作图	转录物的 5′端或 3′端可通过并具末端标记的较长的反义片段与该 RNA 的杂交来鉴别。用 S1 核酸酶处理杂交分子去除单链区，经凝胶电泳分析可测出剩余片段的大小。
引物延伸	聚合酶可延伸引物直达模板末端，然后从模板解离。延伸产物的全长代表着模板的 5′端。
凝胶阻滞	将蛋白质抽提物与标记过的 DNA 片段混合后进行非变性凝胶电泳分析，显示出蛋白质-DNA 复合物的带会延后出现。
DNase Ⅰ 足迹法	用少量 DNase Ⅰ 处理末端标记的 DNA 和蛋白质的混合物，然后进行电泳分析，可显示出与某一 DNA 序列特异结合蛋白的足迹。这一足迹是在切割产物梯状带中的无带区域。
报道基因	为了验证启动子的功能，可将启动子与一个产物易于检测的基因(报道基因）的编码区相连，应在启动子被激活的条件下分析蛋白质产物。
相关主题	DNA 克隆概述（G1） 基因组文库（I1）

组构 cDNA 克隆有一定的组织方式，尤其那些以寡聚 dT 为引物合成的 cDNA 克隆。通常在克隆的一端有一串 A 残基，标志着 3′端。在一串 A 残基上游的一定距离处会有一个终止于终止密码子的可读框（参见 P1)。如果 cDNA 克隆是完整的，会含有一个 ATG 起始密码子，在这之前通常有 20～100nt。真核生物的基因组克隆更大，并含有内含子序列和非转录序列，这些序列增加了了解克隆的组织方式的困难。在分离得到一个 cDNA克隆后，通常要克隆出该基因的基因组序列进行研究。之后的问题就是找出这两个克隆的哪些部分相互对应。这就意味着找出了出现在成熟

的 mRNA 转录物中的基因组序列。cDNA 克隆中不包含的基因组序列通常是内含子、转录起始点的上游以及 3′-加工位点的下游序列（参见 M4 和 O3)。其他要鉴定的重要特征包括转录的起始点、终点和转录调控序列 (参见 M 和 N)。

在基因组 DNA 上定位 cDNA

如果基因组克隆与 cDNA 克隆都有限制图谱，一个重要实验是将酶解后的基因组克隆产物电泳，并用全长或部分 cDNA 为探针进行 Southern 杂交（参见 J1)。用全长 cDNA 克隆为探针可以显示出基因组克隆的哪些限制性酶切片段中会有 cDNA 中的序列。如果存在大内含子，这些片段在限制图谱中可能不相邻。用 cDNA 一端制得的探针可显示出基因组克隆中**该基因的极性**。两类克隆中可能会有一些相同的限制酶切位点，但其间的距离可能不同。这常有利于了解基因组克隆的组织方式。基于这个信息，可对基因组克隆的某些区域进行测序以验证上述结论。

S1 核酸酶作图

这一技术可精确测定 RNA **转录物的 5′端与 3′端**，只是对 5′端与 3′ 端需分别用不同的探针。如图 J4.1 所示 5′端的作图，是将一个末端标记的**反义 DNA** 分子与 **RNA** 样品进行杂交。如果用双链 DNA 来杂交（只有反义链有末端标记)，在杂交缓冲液中应加 80% 甲醛以利于 DNA-RNA 杂交分子比双链 DNA 更易形成。然后用单链特异性的 **S1 核酸酶**处理杂交分子，以移去突出于两端的单链部分，并将样品与分子质量标记或该 DNA 的测序反应样品一起进行 PAGE 分析。再根据放射自显影所显示出的抗核酸酶带的大小来推测 RNA 分子的末端。

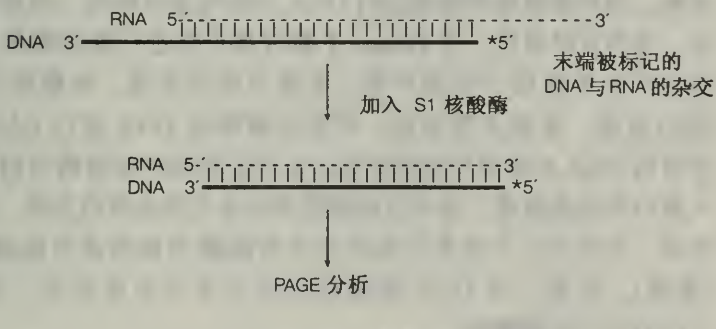


图 J4.1 对 RNA 的 5′端进行 S1 核酸酶作图。* = 末端标记部位

引物延伸

用**反转录酶**沿 **5′端**向 3′端方向延伸反义 DNA **引物**，即从引物与目的序列配对处向聚合酶解离的模板末端延伸，然后将引物延伸的产物与能确定其长度的分子质量标记和（或）不同大小的序列共同电泳，可以测定出 RNA 分子的 5′端（图 J4.2)。

凝胶阻滞

当某一基因的转录产物的 5′端被确定后（如通过 S1 核酸酶作图)，在基因组克隆上的相应位置就是转录起始点。DNA 序列的上游含有控制

基因时空转录的调控序列。转录因子（参见 N1）与特异的调控序列相结合，并有助于转录起始。**凝胶阻滞技术**（**凝胶移位分析**）能显示出与标记核酸**结合蛋白**的阻滞效果，并可被用来检测与调控序列相结合的转录因子。将一段短的标记核酸，如基因组克隆的转录起始点上游区域，与可能会含有结合蛋白的细胞或核的抽提物相混合后，并与对照样品一起进行非变性的琼脂糖或聚丙烯酰胺凝胶电泳分析。在未标记的不同核酸过量存在下，这些核酸与蛋白质非特异结合，而标记分子与蛋白因子特异结合产生一种或多种 DNA 蛋白复合物，经放射自显影在凝胶上作为慢迁移带（**延迟带**）得以显示。

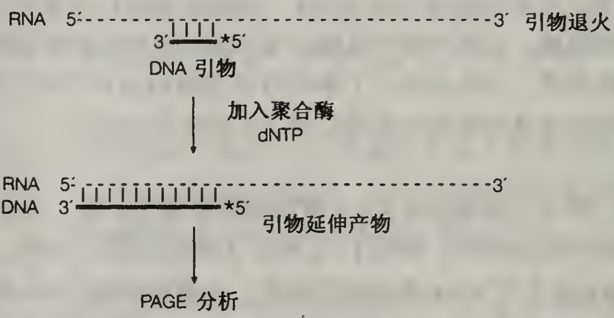


图 J4.2 引物延伸。* = 末端标记部位

DNase Ⅰ 足迹法

虽然凝胶阻滞能显示出蛋白与 DNA 分子的结合，但它并不能提供在所用片段中结合位点的序列，DNase 足迹法可确定与蛋白作用的实际序列区域。同样需要用末端标记的 DNA 片段与蛋白样品（如核抽提物）相混合，待相互结合后，用 **DNase Ⅰ** 酶对复合物进行温和酶解，以期产生平均一个分子被切一次的产物。在蛋白结合区域，核酸酶不能轻易接近 DNA 骨架，很难实现切割。将部分酶解的 DNA 进行 PAGE 电泳分析，在对照 DNA 的泳道中的梯状显示出所有随机的核酸酶切割部位，而在加入蛋白样品的泳道，显示出的梯状带中会有某些带的空缺，或者未切割的区域，正好对应于受蛋白保护免受核酸酶切割的**蛋白结合的 DNA 位点**(**足迹**)。其他一些 DNA 切割试剂也可用于足迹实验，如羟基自由基(·OH)和二甲基磺酸。

报道基因

当控制转录的基因区域（**启动子**，参见 K1）经测序、S1 作图和 DNA-蛋白结合实验确定后，通常要将启动子接上报道基因对其功能进行研究，以验证该启动子所具有的特性。例如将热休克基因的启动子 *HSP70* 与 β-半乳糖苷酶基因的编码区相连，当这一结构体表达，且在有生色底物（X-gal，参见 H1）存在时，就会产生蓝色。若将 *HSP70* 启动子-报道基因的结构体引入一个细胞，并给该细胞或细胞系热激，就会产生 β-gal 转录产物，并可通过检测蓝色来分析蛋白产物。这显示出热激可

以激活通常非活性状态的启动子，其机制是因为一个特定的转录因子结合到启动子的调控序列上，激活该基因（参见 N2）。

J5 克隆基因的诱变

要 点

缺失诱变	从一端开始渐渐缺失 DNA 对于确定特定序列的重要性是很有用的。用可从凹进的 3′端沿 3′向 5′方向移去一条链的外切核酸酶Ⅲ进行单向缺失，再用一种单链特异性核酸酶制成平末端用于连接，经转化产生出具缺失的克隆。
定点诱变	在一个特定位点改变一个或几个核苷酸通常需要将诱变引物与模板退火，再用 DNA 聚合酶合成互补链。过去多用由 M13 制备的单链模板，但现在倾向于用 PCR 技术。
PCR 诱变	用正向与反向突变的引物及可结合常用载体序列的另一对引物，进行两组 PCR 反应来分别扩增待突变的 DNA 的 5′与 3′部分。然后将这两组 PCR 产物混合，并进行只用外部引物的另一 PCR。这些产物中部分再通过亚克隆代替最初分子中要突变的区域。
相关主题	DNA 克隆概述（G1） 噬菌体载体（H2） 连接、转化与重组体分析（G4） 聚合酶链反应（J3）

缺失诱变

对于小基因组或是生活周期短的生物体，有可能在这些生物体内制作突变体并进行解析，但对于像人这样的生物体，不易进行这种操作或是不可接受的。但若克隆基因已被分离，相当容易在体外对它们进行突变，并通过体外或体内表达该突变基因来分析突变效果。对 cDNA 克隆与基因组克隆制作缺失诱变是可行的。对于 cDNA 克隆，通常从编码区的一端逐渐删除，产生 N 端或 C 端**截短的蛋白**（表达后），从而可发现整个蛋白的哪一部分（或结构域）具有特殊特性。例如某一特定蛋白的 N 端区域可能是 DNA 结合域，中间部分是 ATP 结合位点，而 C 端区域则有助于蛋白质相互作用形成二聚体。对于基因组克隆，鉴别出转录起始点后，逐步删除其上游的序列以发现具有启动子和调控功能的最短上游序列。虽然只要其作用位点落在可操作的部位，可用限制酶制作缺失突变体，而用外切酶制作**单向缺失**的方法也很常用。

图 J5.1 显示克隆进质粒载体多克隆位点（MCS）的 cDNA 在插入片段两端的多克隆位点处存在几个限制性酶切位点，选择两个对载体和插入片段单一的，并分别可给出 5′和 3′凹端的酶切位点来进行单向缺失体的制作。**外切核酸酶Ⅲ**可以从凹进的 3′端沿 3′→5′方向移去一条核苷酸链，而并不从 3′凸端。因此在图 J5.1 中外切核酸酶Ⅲ只能随时间逐步移去插

入片段的下链。用 **S1 或绿豆核酸酶**（参见 G1）等对等分样品进行不同时间处理，去掉两端的突出单链部分形成平末端，然后经连接重新环化成质粒，再进行转化（参见 G4），将产生一系列缺失不同长度的亚克隆。这一技术常被用来获取克隆的全序列，因为可用同一测序引物对一系列相差 200bp 左右的克隆进行序列分析。大约三分之一的缺失体仍保持正确的可读框，能用来表达截短蛋白。可用同样方法对 cDNA 另一端进行缺失。

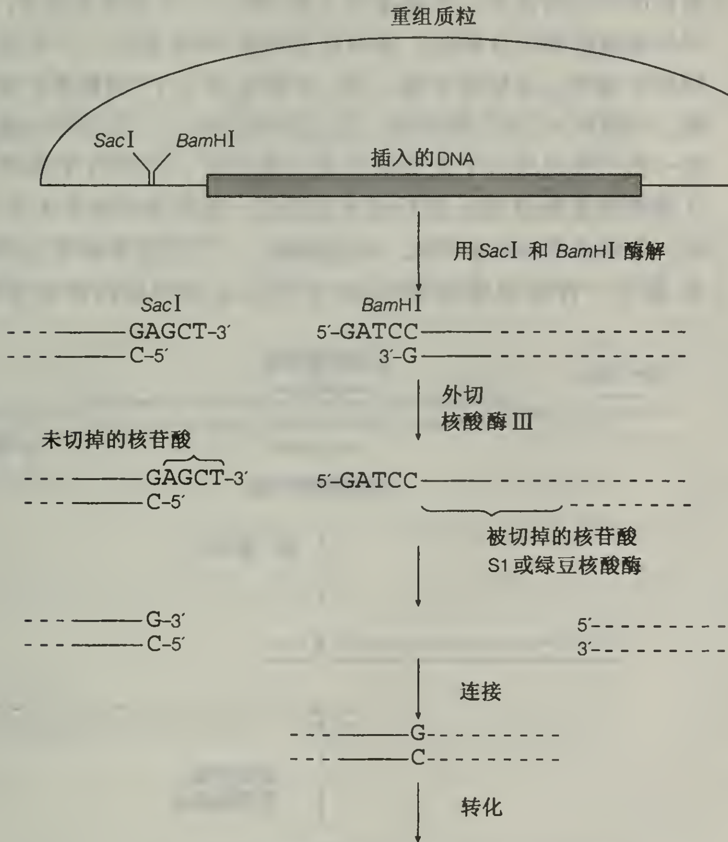


图 J5.1　单向缺失诱变

定点诱变

仅改变一个序列中的一个或几个特定核苷酸对于验证某一假说是非常有效的。转录因子结合位点的每一残基的重要性可通过将各残基逐个突变来检验。对蛋白质中可能起重要作用的氨基酸可通过改变其对应的 cDNA 的核苷酸来改变该蛋白的组成，进而对所产生的突变蛋白进行功能分析以检测突变效果。最初的定点诱变是用**单链模板**（由 M13 亚克隆制备）及含有待突变的**寡聚核苷酸引物**来进行的。引物与模板退火后，用 DNA 聚合酶延伸，用 DNA 连接酶缝合切口，并将不完全配对的双链转化细菌。某些细菌能移去错配，产生出所期望的点突变，当然也会有一些含原初序列的克隆产生。因为这种方法效率低（并非所有克隆都是突变体），且经

M13 的亚克隆来制备单链 DNA 很费时间，现在多用 PCR 来进行定点诱变。

PCR 诱变

用 PCR 方法（参见 J3）来制作缺失和点突变有好几种方法，图 J5.2 显示了进行点突变的一种方法。设计一对已含有突变序列的、且彼此重叠至少 20nt 的引物。如果待突变的 DNA 是被克隆在标准载体上，载体上会有标准的引物位点，如 pGEM 上的 SP6 与 T7 启动子位点，这些标准引物可与**诱变引物**组合使用。分别进行两次 PCR 反应，一次用 SP6 与反向引物来扩增插入片段的 5′端，另一次用正向与 T7 引物来扩增插入片段的 3′端。将两次 PCR 产物纯化、混合并用 SP6 与 T7 引物一起来扩增，结果惟一的产物应是一个全长的突变了的分子。只有当 5′端有义链与 3′端反义链借助重叠的 20nt 进行退火或反之，然后进行延伸才会产生上述产物。这一方法几乎 100%有效，而且很快。一个更方便的可选择的版本，至少是 基于一种商用诱变试剂盒，包括用正向和反向诱变引物（图J5.2）来

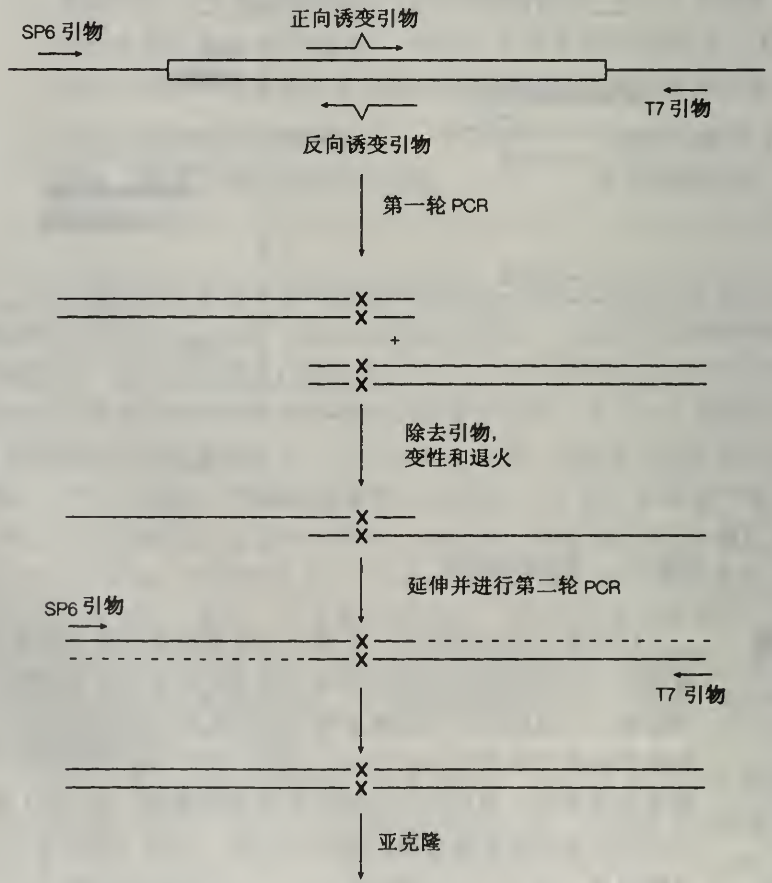


图 J5.2 PCR 诱变。X 表示 PCR 产物中的突变位点

延伸含有目的基因的整个质粒。在反应结束时，一些产物将含有在两条链中都存在突变的环状分子及在每一引物序列末端均存在切口。这些分子能直接用于转化产生含有突变质粒的克隆，而无需用限制性内切核酸酶消化及连接。因为所有突变体在被使用前要经测序进行检查，因此最好不用 PCR 形成的 DNA 整个分子，而将含有突变区的较小片段亚克隆进原初克隆的相同位点，这样只有被转移区域才需要测序检查。

J6 克隆技术的应用

要 点

应 用	基因克隆技术有多种应用，包括重组蛋白的生产、遗传修饰生物体的构建、DNA 指纹分析、诊断试剂盒以及基因治疗。
重组蛋白	将编码珍贵蛋白的基因插入质粒并在细菌中表达出大量生产重组蛋白。如果翻译后的修饰很重要，则基因须在真核细胞中表达。
遗传修饰生物体	向一个生物体引入一个可在其内增殖的外源基因，构成一种遗传修饰生物体。例如已做出的转基因羊能在其奶汁中生产外源蛋白。
DNA 指纹分析	将基因组 DNA 与可识别单核苷酸重复的探针进行 Southern 杂交，产生出为某一个体所独有的图谱，可被用来进行指纹分析。这在法医学、动植物育种以及进化研究中有着广泛应用。
医学诊断	可根据已克隆的具有重要医学价值基因的序列信息设计多种诊断试剂盒，来对多种疾病进行预测与确诊。
基因治疗	通过向病人导入特定基因试图来治疗某种遗传疾病的方法称为基因治疗。

相关主题

基因组复杂度（D4） 质粒载体的设计（H1）
真核生物载体（H4） 癌基因的分类（S2）

应用

基因克隆技术促进了生物学研究的快速发展，并在日常生活的多个领域有着越来越广泛的应用。这些应用包括药用蛋白与非药用蛋白的生物技术生产、基因修饰生物体的诞生（尤其是食物生产的改良）、医学诊断试剂盒的开发、PCR 及克隆技术在法医和进化研究中的应用以及用基因治疗医治遗传病的尝试等。本节将介绍这些应用中的一部分。

重组蛋白

已知许多正常产量很少的蛋白在多种疾病患者体内产生缺失或缺陷，包括生长激素、与糖尿病相关的胰岛素、与某些免疫疾病相关的白介素以及血友病相关的凝血因子Ⅷ。在基因克隆及用重组 DNA 技术进行蛋白生产之前，须从动物组织或捐献的人血中纯化这些分子。这两种来源都有欠缺，如来自动物细胞的蛋白有轻微的功能差异以及血液中可能的病毒污染（如 HIV、CJD）。用特定的非病原性的生物体来进行蛋白质的基因工程生

产避免了这些问题，所以制药公司和生物技术公司致力于这项技术的发展。最初用细菌来生产蛋白是惟一常用方法，且使用不含内含子的cDNA。cDNA 须与原核生物的转录和翻译信号相连，并插入多拷贝质粒（参见 H1）。但占总细胞蛋白 30%的过量表达蛋白经常会发生沉淀或不溶，也缺乏真核生物那样的翻译后修饰（参见 Q4）。有时采用融合表达来解决上述这些问题。融合蛋白经切割可释放出目的蛋白。但随后采用真核细胞（酵母或哺乳动物细胞）进行蛋白生产的方法更方便可行。可支配血友病的人凝血因子Ⅷ蛋白可在转染人基因组 186kb 片段的仓鼠细胞系中生产。这样该细胞系已被遗传修饰，而现在已可对所有生物体进行**遗传修饰**。重组蛋白也可通过诱变（参见 J5）引入新的氨基酸来进行修饰。这可对蛋白进行改良，如构建更稳定的酶应用于洗衣粉等。

遗传修饰生物体

将克隆基因引入能够形成整个机体组织的细胞可构建成**遗传修饰生物体**（genetically modified organisms，GMO）。在真核生物中，如果被引入的基因来自另一生物体，形成的转基因植物或动物可以由常规育种来繁殖。已构建成几种转基因植物，并已对其食物生产的安全性进行了检验。GMO 的例子之一是一种促成熟酶基因失活的西红柿。由于该酶的缺失这一品系的西红柿变软所需的时间更长，因此也就能储存更久，同时还有其他一些改良特性。转基因羊已获成功，并希望在其乳汁中生产有价值的蛋白。目的基因需要连接上一个羊的启动子以保证在乳腺中表达，从乳汁中纯化蛋白比从培养细胞或血液中纯化更容易。一种转基因生物体的定义是指含有一个外源基因的生物体，但该术语现在常被用来指生物体经过遗传操作后含有多个外源基因、一种内源基因的多拷贝或一个基因被破坏（**基因剔除**）。该术语不常用在一些动物克隆中（产生同一个体）常见的增加外源基因的最极端形式，即其中全部的基因都被另一个细胞核来源的基因取代。这种用一个成熟细胞核取代一个卵子细胞核的方案（**核移殖**），在 1997 年曾用来创造了绵羊多利（Dolly）。这个著名的动物克隆的例子造成了很多争论，因为它使从成熟细胞克隆人的可能性大大提高。许多国家现在已经立法禁止多种形式的人的克隆，这将阻止用于治疗的替代细胞和（或）器官的开发。同卵双生的双胞胎是自然形成的人的克隆。

DNA 指纹分析

基因组测序计划已从人类基因组中鉴别出许多重复序列（参见 D4），其中一些是简单重复序列，不同个体间重复数目不同但可遗传（VNTR）。如果将一个家族成员的经限制酶酶解后的基因组 DNA 与能探测某种重复序列的探针进行 Southern 杂交，每一样品都会出现一套不同长度的带（在两个限制酶切位点间的重复长度不同）。一个杂交基因座（成对等位基因）展示于图 J6.1，其中一些带与母本的相同，另一些与父本的相同，且这些带型将与无关个体不同。个体之间在每一简单重复位点上的图谱不同意味着用少数探针就可使两个体间形成相同图谱的可能性几乎为零。这

就是 **DNA 指纹技术**，被用于法医学来判别无罪和证实有罪。同时也在人类的家谱检测，商用动物的血统检测以及野生动物的交配习性的研究方面有着广泛应用。图 J6.1 也显示 DNA 指纹图谱怎样在少量的样品中实现，例如在某犯罪现场留下的一个血斑或发囊。可以根据重复的独特的侧翼序列设计一对 PCR 引物来替代用限制性内切核酸酶 E 在每一个 VNTR 末端消化 DNA 以及 Southern 杂交（在图 J6.1 中用箭头表示），该 VNTR 可被扩增并通过凝胶电泳后染色可直接观察。

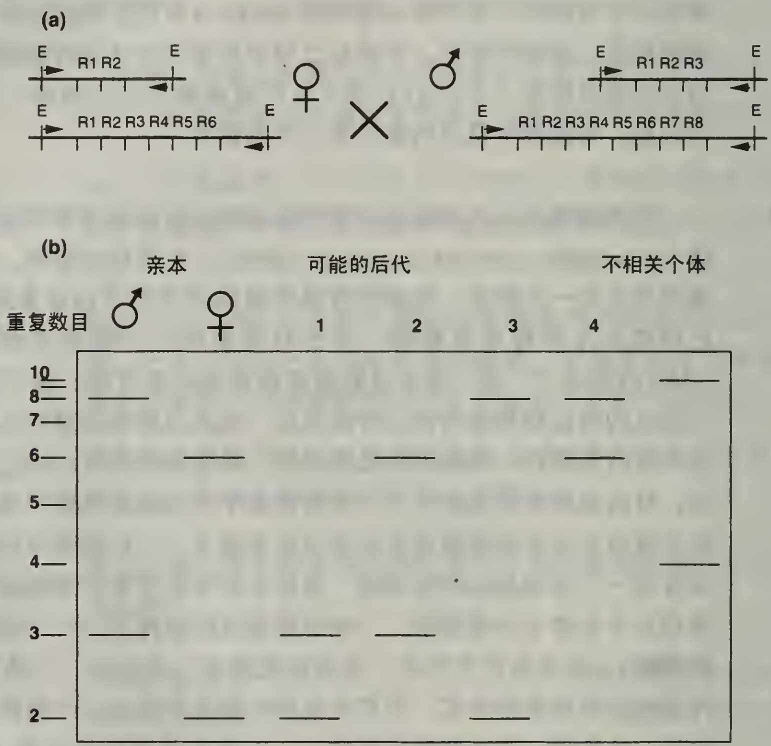


图 J6.1 DNA 指纹分析显示两个 VNTR 等位基因可能的遗传方式（见正文）。(a) 亲本 VNTR 等位基因；(b) VNTR 等位基因的琼脂糖凝胶分析

医学诊断

许多疾病起因于变异。诸如肌肉萎缩或囊性纤维化这样的遗传病，个体出生时即带有引起该症状的错误基因。许多癌症的产生是由于体细胞中控制细胞生长的基因发生自发突变的结果（参见 S2）。有关遗传病与癌症基因的克隆显示出某些突变较为常见，且其中一些与更为严重的疾病相关。利用序列信息设计 PCR 引物和探针，来开发多种检测方法用以对病人进行临床上重要突变的检查。利用这些检测方法，若发现父母均是某一突变的杂合体，可被告之未出生的孩子是否会得如肌肉萎缩或囊性纤维化这样的遗传病（从父母各继承一个错误基因），进而父母可以考虑是否要这个孩子。检验基因中突变的存在可证实基于其他临床症状所作出的诊断。对于癌症，了解哪个致癌基因发生了突变，以什么方式发生突变，将

有助于决定最好的治疗程序以及为开发新疗法提供信息。

基因治疗

人类已在尝试通过向病人导入缺陷基因的正常拷贝来治疗遗传病，这就是基因治疗。用于体内基因治疗的基因可被自身支配或被克隆进用作载体的、能复制但不会致病的缺陷型病毒内（参见 H4）；对某些疾病，患者骨髓被破坏，可通过引入（***ex vivo* 基因治疗**）携有正常基因或保护基因（如对核酶，参见 O2）的细胞来进行治疗。基因治疗会在病人中产生转基因体细胞（见上），基因治疗虽不很成熟，但它的潜力巨大。

（程　凌　译　王薛林　校）

K1 转录的基本原则

要 点

转录概述	转录是指以双链 DNA 为模板合成单链 RNA。RNA 的合成沿 5′→3′方向进行，其序列与 DNA 有义链相一致。
起 始	RNA 聚合酶是负责转录的酶，它结合到被称为启动子的特定 DNA 序列上以起始 RNA 的合成。这些序列位于蛋白质编码区的上游（5′端），含有一些短而保守的 DNA 序列，在不同启动子中这些序列是相同的。RNA 聚合酶结合到双链 DNA 的启动子序列上，引起 DNA 局部解旋。合成第一个 RNA 碱基的位置称为起始位点，其位置称为 +1。
延 伸	RNA 聚合酶沿着 DNA 链移动，顺次合成 RNA 链。DNA 在移动的聚合酶抵达之前解旋，在其经过之后又恢复成双螺旋。
终 止	RNA 聚合酶识别终止子，导致不再有新的核苷酸插入，该序列通常是发夹结构。一些终止子需要一种称作 ρ 的辅助因子来终止转录。
相关主题	核酸结构（C1） 大肠杆菌 RNA 聚合酶（K2） 大肠杆菌 σ^{70}启动子（K3） 转录的起始、延伸与终止（K4） 色氨酸操纵子（L2）

转录概述

转录就是以 DNA 为模板的 RNA 的酶促合成。这是基因表达全过程的第一步，并最终导致由基因所编码蛋白的合成。转录由 **RNA 聚合酶**催化，它需要双链 DNA 模板与前体核糖核苷酸 ATP、GTP、CTP 和 UTP（图 K1.1）。RNA 合成的方向总是固定的，即从 RNA 分子的 5′→3′（参见 C1）。通常两条 DNA 链中只有一条会转录成 RNA，其中一条链称为**有义链**，RNA 的序列就是有义链上脱氧核苷酸序列的直接拷贝（以 U 替代 T）。另一条链称为**反义链**，也称**模板链**，因为它用作 RNA 合成过程中核糖核苷酸碱基配对的模板。

起 始

转录的起始涉及到 RNA 聚合酶与双链 DNA 的结合。RNA 聚合酶通常是多亚基酶。它们结合到双链 DNA 并在被称作**启动子**的部位起始转录（图 K1.2）。启动子是基因起始处的 DNA 序列，即在编码区的 5′端（**上游**）。不同基因的启动子序列元件经常是保守的。不同基因间启动子的差别会导致转录起始效率的差异且与其调控有关（参见 L）。启动子内的短

的保守序列是聚合酶或其他 DNA 结合蛋白的结合位点以起始或调控转录。

为使模板链能与碱基配对，DNA 双螺旋必须局部解旋。解旋从 RNA 聚合酶结合的启动子部位开始。然后聚合酶从一特定的、被称为**起始位点**的核苷酸处起始 RNA 链的合成。这一位点被定义为基因序列的 +1 位置(图 K1.2)。结合到 DNA 模板上的 RNA 聚合酶及其辅助因子通常被称为**转录复合物**。

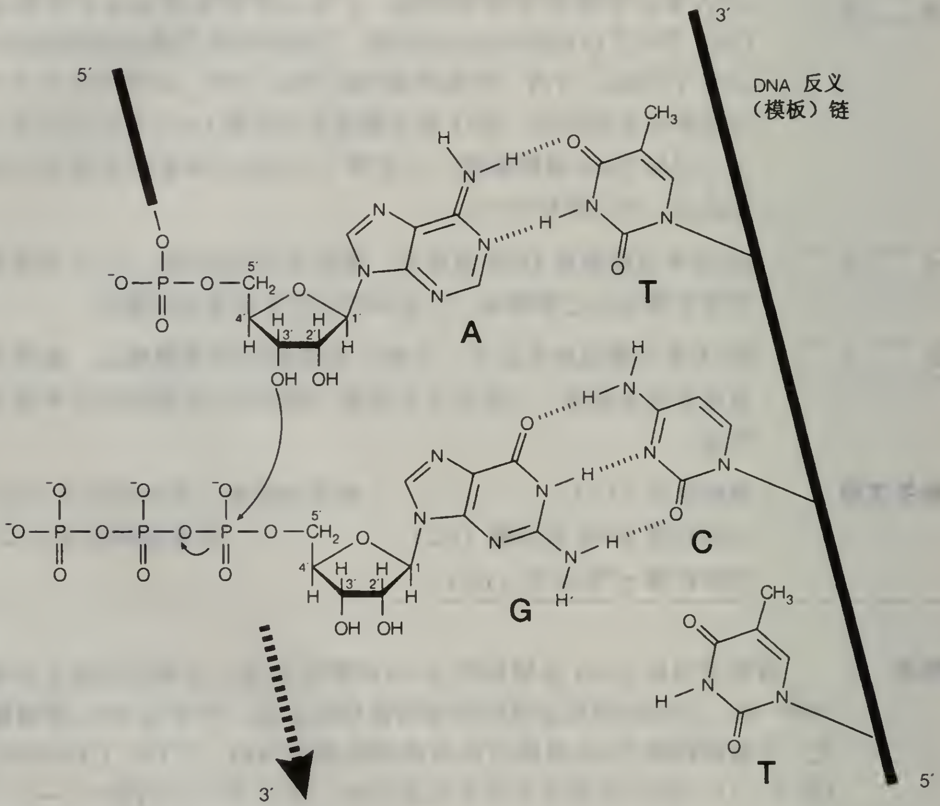


图 K1.1 转录过程中磷酸二酯键的形成

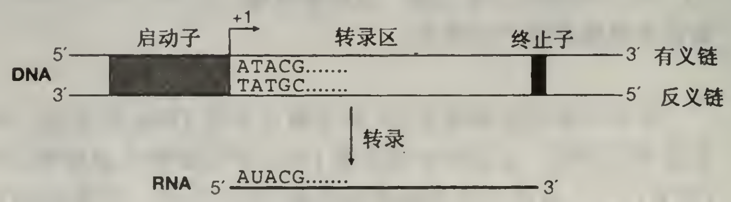
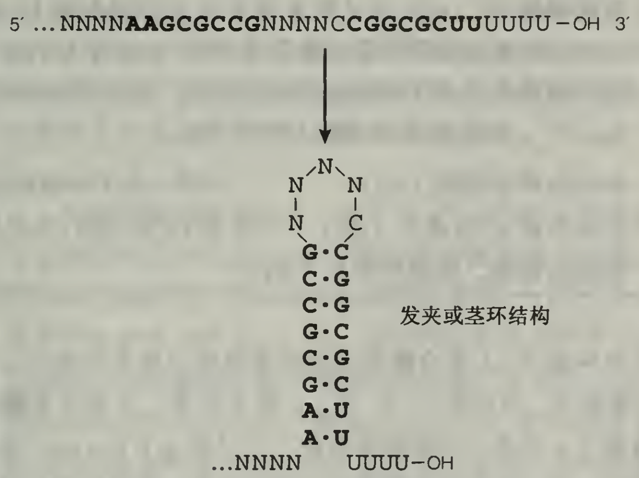


图 K1.2 典型的转录单元结构，标明了启动子、终止子序列和 RNA 产物

延 伸 RNA聚合酶向正在增长的RNA链的3′端共价地添加核糖核苷酸（图K1.1）。因此聚合酶是沿5′→3′方向延伸正在增长的RNA链。同时酶自身沿着反义DNA链（模板）的3′→5′方向移动。酶移动时局部解旋DNA，使两条DNA链分开暴露出模板链，供核糖核苷酸碱基配对并向增长着的RNA链的3′端共价添加核糖核苷酸。在聚合酶经过之后又重新形成双螺旋。37℃时，*E. coli* RNA聚合酶完成该反应的速度是每秒约40个碱基。

终 止 转录终止即转录复合物的解体和RNA合成的终止，发生在特定的DNA序列即**终止子**处（参见图K1.2和K2、K3）。这些序列经常含有一些自身互补区，能在RNA产物上形成**茎环**或**发夹**二级结构（图K1.3）。这些结构使聚合酶停顿并随即终止转录。

一些终止子可以不需要辅助因子而独立终止转录，另一些则需要rho蛋白（ρ）作辅助因子。在终止反应中，RNA-DNA杂交体分开以重新形成双链DNA，同时RNA聚合酶和所合成的RNA从DNA链上被释放下来。



图K1.3 RNA发夹结构

K2 大肠杆菌 RNA 聚合酶

要 点

大肠杆菌 RNA 聚合酶	RNA 聚合酶负责 RNA 的合成（转录），其核心酶由 2α、1β、1β′和 1ω 亚基组成，负责转录延伸。σ 因子对于转录的正确起始是必需的。因此核心酶加上 σ 因子所组成的完整的酶称为全酶。
α 亚基	RNA 聚合酶中有两个 α 亚基，它们可能与启动子的结合有关。
β 亚基	RNA 聚合酶中有一个 β 亚基，抗生素利福平和利迪链菌素都可与 β 亚基结合。β 亚基可能与转录的起始和延伸有关。
β′ 亚基	RNA 聚合酶中有一个 β′亚基，它可能与模板 DNA 的结合有关。肝素可与 β′亚基结合。
σ 因子	σ 因子是核心酶以外的单独成分。大肠杆菌编码几种 σ 因子，最常见的是 σ^{70}。σ 因子对在正确启动子位点起始转录是必需的。这一功能是通过降低核心酶对非特异 DNA 序列的结合，同时增强对特异启动子序列的结合来实现的。转录产物的长度达到 8～9nt 时，σ 因子就从核心酶上释放下来。
相关主题	转录的基本原则（K1）　不同 σ 因子对转录的调节（L3） 大肠杆菌 σ^{70}启动子（K3）　三种 RNA 聚合酶:性质与功能(M1) 转录的起始、延伸与终止（K4）

大肠杆菌 RNA 聚合酶

大肠杆菌 RNA 聚合酶是大肠杆菌细胞中最大的酶之一，该酶至少由 5 个亚基组成，分别是 α、β、β′、ω 和 σ 亚基。在称为**全酶**的完整的酶分子中有两个 α 亚基，另四种亚基各一个（即 $\alpha_2\beta\beta'\omega\sigma$）。全酶是转录起始所必需的。但 σ 因子对于转录延伸并不必需，转录起始后就从转录复合物上释放下来。剩下的酶沿 DNA 链移动，称为**核心酶**，其结构是 $\alpha_2\beta\beta'\omega$。大肠杆菌 RNA 聚合酶在 37℃ 时合成 RNA 的速度是每秒约 40nt，需有 Mg^{2+} 激活。该酶是非球状结构，在一圆柱形孔道旁有一突起。孔道的大小表明它可以直接结合 16 bp 的 DNA。整个聚合酶所结合的 DNA 区域覆盖了约 60 bp。

虽然多数 RNA 聚合酶都像大肠杆菌 RNA 聚合酶一样有多亚基结构，但必须注意的是多亚基结构并非必需的。噬菌体 T3 和 T7 所编码的 RNA 聚合酶（参见 H1）只有一条多肽链，比细菌的多亚基酶小得多。它们能快速合成 RNA（37℃ 时每秒 200 nt）并识别自身特异的 DNA 结合序列。

α 亚基

核心 RNA 聚合酶有两个相同的 α 亚基，该亚基由 *rpoA* 基因编码。α 亚基对于核心酶的组装是必需的，但尚未发现它有明确的转录功能。噬菌体 T4 感染大肠相菌时，α 亚基被修饰，其中的一个精氨酸被腺苷二磷酸（ADP）核糖基化。此时 RNA 酶结合启动子的能力降低，表明 α 亚基可能在启动子识别中发挥一定作用。

β 亚基

核心酶中有一个 β 亚基，这一亚基被认为是 RNA 聚合酶的催化中心。这一观点的强有力证据来自对抑制 RNA 聚合酶转录的抗生素的研究。重要的抗生素**利福平**是 RNA 聚合酶的强抑制剂，它能阻止起始但不影响延伸。这一类的抗生素并不抑制真核细胞的 RNA 聚合酶，因此在医学上被用于治疗革兰氏阳性菌的感染和结核病。已证明利福平是与 β 亚基结合的。能导致对利福平抗性的突变被定位在 *rpo B* 基因内，该基因编码 β 亚基。另一种抗生素**利迪链菌素**能抑制转录延伸，能产生对该抗生素抗性的基因突变也被定位在 *rpoB* 基因内。这些研究表明 β 亚基可能含有两个结构域，分别负责转录的起始和延伸。

β′ 亚基

核心酶中有一个 β′ 亚基，由 *rpo C* 基因编码。这一亚基结合了两个 Zn^{2+}，这两个 Zn^{2+} 被认为参与了该酶的催化功能。多阴离子**肝素**能结合到 β′ 亚基上。在体外肝素能抑制转录，并与 DNA 竞争结合到聚合酶上。这表明 β′ 亚基可能负责酶与模板 DNA 结合。

σ 因子

大肠杆菌中最常见的 σ 因子是 σ^{70}（因其分子质量为 70kDa 而得名）。与 σ 因子的结合使 RNA 聚合酶由核心酶转变为全酶。σ 因子在启动子识别中起关键作用，但对转录延伸并不需要。σ 因子是通过将核心酶对非特异序列的亲和力降低 10^4 倍，同时增加其对启动子的亲和力来帮助启动子识别的。许多原核生物（包括大肠杆菌）有多种 σ 因子。它们与识别各种类型的启动子序列有关（参见 L3）。RNA 链延伸至 8～9nt 时，σ 因子会从 RNA 聚合酶中释放出来。然后核心酶沿 DNA 模板移动并合成不断增长的 RNA 链。释放出来的 σ 因子可以再与其他核心酶结合重新起始转录。细胞内 σ 因子数量只有核心酶的 30%。因此在任何时刻只有 1/3 的聚合酶能以全酶的形式存在。

K3 大肠杆菌 σ^{70}启动子

要 点	
启动子序列	启动子含有 RNA 聚合酶特异性结合和转录起始所需的保守序列。
启动子大小	启动子区域长约 40 bp。在这段序列内，有一些对于启动子的功能很关键的高度保守的短序列。
-10 序列	-10 序列是几乎所有启动子都含有的一个 6 bp 的序列。该六聚体通常位于转录起点上游 10bp 处。共有的 -10 序列是 TATAAT。
-35 序列	-35 序列是另一个在大多数启动子中都可找到的 6 bp 序列。该六聚体一般位于转录起始点上游 35bp 处。共有的 -35 序列是 TTGACA。
转录起始点	转录起始点处的碱基多为嘌呤，G 比 A 更常见。
启动子效率	不同启动子序列与不同基因的转录速率都千差万别。可调控的启动子（如 *lac* 启动子）要结合辅助性的激活因子如 cAMP 受体蛋白（CRP）才能被激活。保守启动子序列的不同类型（如热休克启动子）只能被含有不同 σ 因子的 RNA 聚合酶所识别。
相关主题	克隆基因的组构（J4） 转录的基本原则（K1） 大肠杆菌 RNA 聚合酶（K2） 转录的起始、延伸与终止（K4） 乳糖操纵子（L1） 不同 σ 因子对转录的调节（L3）

启动子序列　RNA 聚合酶结合到待转录序列上游的特定起始位点，即启动子部位。尽管识别不同的启动子需要不同的与 RNA 核心酶结合的 σ 因子，但大肠杆菌中最常见的 σ 因子是 σ^{70}。启动子的特性最初是通过能增加或降低基因转录速率的突变来鉴定的，例如 *lac* 操纵子（参见 L1）。启动子位于通常被称为 +1 位点的转录起始位点的上游（参见 K1）。按照这一编号方法，启动子序列编号为负数，反映其在转录起始位点上游的距离。大肠杆菌启动子的突变分析表明只有很短的几个保守序列是启动子行使功能所必需的。

启动子大小　σ^{70}启动子由 40 至 60 bp 的序列组成。-55 到 +20 之间的区域都被聚合酶的结合，-20 到 +20 之间的区域被强有力地保护起来，不受 DNase Ⅰ 消化（参见 J4）。这表明该区域与聚合酶紧密地结合从而阻止了

核酸酶接近 DNA。启动子序列的突变分析表明直至 -40 位置的区域对于启动子功能仍是必需的。在 -10 和 -35 位置的两个 6 bp 序列对于大肠杆菌中启动子的功能尤其重要。

-10 序列

大肠杆菌的 σ^{70}启动子中最保守的序列是一个在许多不同的大肠杆菌基因启动子中都存在的 6 bp 序列。这一序列的中心位于相对于转录起始位点的 -10 位置（图 K3.1)。有时也称这一序列为 **Pribnow 框**，因为是由 Pribnow 于 1975 年发现的。它有一段共有序列 **TATAAT**，在比较了许多启动子序列后发现这段共有序列是在各个位置上最常出现的核苷酸。头两个碱基（TA）和最后的 T 是最保守的。该六聚体距转录起始位点 5 至 8 bp。其间的间隔序列是不保守的，但距离很重要。 -10 序列似乎是聚合酶启动 DNA 解旋的序列（参见 K4)。

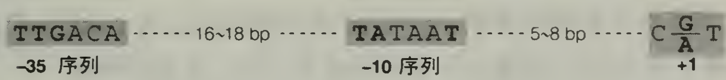


图 K3.1 *E. coli* 启动子的共有序列（最保守的序列用黑体表示）

-35 序列

在 -35 位置附近的上游区也有一段保守的六聚体序列（图 K3.1)。其共有序列是 TTGACA，该序列在高效启动子中是非常保守的。该六聚体的前三个碱基最保守的，在 90%的启动子中该序列距 -10 框有 16 至 18 bp。在这两个保守元件之间的间隔序列并不重要。

转录起始点

90%的基因的转录起始位点是嘌呤（图 K3.1)，G 比 A 更常见。通常在起始点核苷酸的两侧分别是 C 和 T 碱基（如 CGT 或 CAT)。

启动子效率

上面提到的序列是强启动子中典型的共有序列，但不同启动子的序列之间有很大差异，而且其转录效率相差可高达 1000 倍。总的说来不同启动子区域的功能如下：

- -35 序列组成增强与聚合酶 σ 因子相识别和相互作用的识别区。
- -10 序列对于 DNA 解旋很重要。
- 起始位点附近的序列可影响转录起始。

最初转录的 30 个碱基序列也会影响转录。该序列控制 RNA 聚合酶离开启动子，使另一聚合酶复合物得以重新起始转录的速率，由此影响转录的速率乃至整个启动子的强度。DNA 模板的负超螺旋可增强转录起始，这可能是因为负超螺旋结构需要较少的能量来解旋 DNA，由此可见起始反应中两链分开的重要性。一些启动子序列与通常条件下可被较强转录的共有序列不十分相似。一个例子是 ***lac* 启动子 P_{lac}**，它需要一个辅助激活因子 **cAMP 受体蛋白**（CRP）结合到 DNA 上一个靠近启动子的位点，以

增强聚合酶的结合和转录起始（参见 L1）。其他启动子如与热休克相关基因的启动子，含有只能被与一般 σ^{70} 的不同 σ 因子结合的 RNA 聚合酶所识别的不同的共有启动子序列（参见 L3）。

K4 转录的起始、延伸与终止

要 点

启动子结合 σ因子可增强核心 RNA 聚合酶 $\alpha_2\beta\beta'\omega$ 与启动子结合的特异性。聚合酶通过沿着 DNA 链滑动并与启动子 DNA 形成闭合复合物来找到启动子的 -35 序列和 -10 序列。

DNA 解旋 在聚合酶的作用下大约有 17 bp 的 DNA 解旋，形成一个开放的复合物。许多启动子中负超螺旋 DNA 可增强 DNA 解旋，但 DNA 促旋酶亚基基因的启动子却被负超螺旋所抑制。

RNA 链起始 RNA 的合成无需引物，最初的 9 个核苷酸是在聚合酶不沿 DNA 链移动或者 σ 因子存在的情况下合成的。RNA 聚合酶首先要经过多次无效的链起始，在成功起始之后 σ 因子释放出来，形成一个负责 RNA 链延伸的四聚体复合物。

RNA 链延伸 RNA 聚合酶沿 DNA 链移动，DNA 链上始终保持一段称为转录泡的解旋区。在转录泡内的 RNA 5′端的 10～12 个核苷酸始终与 DNA 模板链配对。聚合酶在转录泡的前方解旋 DNA，而在其后方复旋 DNA。

RNA 链终止 基因 3′端自身互补的序列会在 RNA 中形成发夹结构作为终止子。发夹的茎部通常有较高的 G-C 含量，使其有很高的稳定性，从而使聚合酶在此停顿。发夹之后通常是 4 个或更多的 U 碱基，导致 RNA 和反义 DNA 链的弱结合。这利于 RNA 链的解离，从而终止转录。

依赖 ρ 的转录终止 有些基因的终止序列需要一个辅助的蛋白因子 ρ（rho）才能有效地终止转录。ρ 结合到单链 RNA 的特定位点上，它水解 ATP 并沿 RNA 链向转录复合物移动，与之结合后使聚合酶终止转录。

相关主题

DNA 超螺旋（C4）
DNA 复制：概述（E1）
转录的基本原则（K1）
大肠杆菌 RNA 聚合酶（K2）
大肠杆菌 σ^{70}启动子（K3）
原核生物的转录调控（L）
真核生物的转录（M）

启动子结合 RNA 聚合酶的核心酶 $\alpha_2\beta\beta'\omega$ 对 DNA 有一种非特异性的亲和力，这被称为**松散结合**并相当稳定。当 σ 因子加入核心酶形成全酶之后，全酶对非特异 DNA 的亲和力显著降低达 20 000 倍；同时 σ 因子使全酶结合到正确的启动子位点的能力增强 100 倍。总的效果就是显著地增强了全酶与正

确的启动子结合位点相结合的特异性。在大肠杆菌基因组中全酶发现并与启动子结合是很迅速的。这一过程太快了，因此不可能是通过与 DNA 的反复结合与解离来实现的，一般认为聚合酶沿着 DNA 链滑动直达启动子序列处。在启动子处，聚合酶识别双链的 −35 序列和 −10 序列。聚合酶与处于碱基配对状态的启动子 DNA 所形成的最初的复合物被称为**闭合复合物**。

DNA 解旋

为使反义链实现碱基配对，DNA 双螺旋必须被聚合酶解旋。负超螺旋能增进许多基因的转录，因为它有利于聚合酶的解旋。但是也有一些启动子不能被负超螺旋激活，说明天然 DNA 拓扑构象的区别可能影响转录，这可能是因为双螺旋上 −35 序列和 −10 序列的空间关系的差异。例如，DNA 促旋酶亚基基因的启动子就被负超螺旋所抑制。DNA 促旋酶负责大肠杆菌基因组的负超螺旋的形成（参见 C4），因此这可能是 DNA 促旋酶蛋白表达中的一个精细的反馈调节。DNA 的初始解旋导致 DNA 与聚合酶形成**开放复合物**，这一过程被称为**紧密结合**。

RNA 链起始

几乎所有 RNA 起始位点都由一个嘌呤残基组成，其中 G 比 A 更常见。与 DNA 合成不同（参见 E），RNA 合成不需引物（图 K4.1）。链以 GTP 或 ATP 开始，接着继续链的其余部分合成。聚合酶最先掺入最初两个核苷酸，并形成一个磷酸二酯键。最初的九个碱基依次加上，酶并不沿 DNA 链移动。这九个核苷酸每加上一个，这条链都很有可能被流产。这种无效起始对于转录的配体速率很重要，因为它对于聚合酶要花多长时间离开启动子并允许另一个聚合酶来起始新一轮转录有重要作用，启动子清除的最短时间是 1～2s，相对于转录的其他步骤，这是一个较长的过程。

RNA 链延伸

起始成功之后，聚合酶释放出 σ 因子，形成聚合酶-DNA-新生 RNA 的**三聚复合物**，使聚合酶沿 DNA 链移动（**启动子清除**），允许另一 RNA 聚合酶全酶从启动子处再度起始转录。DNA 的解旋区（称为**转录泡**）可能和聚合酶一起沿 DNA 链移动。解旋 DNA 区域的大小稳定地保持在约 17 bp（图 K4.2），RNA 的 5′端与反义 DNA 链形成约 12 bp 的杂合双螺旋。这正好小于 RNA-DNA 双螺旋的一圈。大肠杆菌聚合酶以每秒 40 个核苷酸的速度移动，但对于不同的局部 DNA 序列速度会有所差异。保持一个短的 DNA 解旋区表明聚合酶在转录泡之前解旋 DNA，而在其之后复旋 DNA。每加一个核苷酸到 RNA 链上 RNA-DNA 双螺旋就要旋转一次。

RNA 链终止

RNA 聚合酶在到达转录单位末尾的**终止序列（终止信号）**之前一直结合在 DNA 上并持续转录（图 K4.3）。最常见的终止信号是一个 **RNA 发夹**结构，其中的 RNA 转录物是自身互补的。因此这种 RNA 能形成一个稳定的、带一个茎和一个环的发夹结构。通常茎部结构富含 G-C，利于碱

基配对的稳定性（G-C 碱基配对比 A-U 配对稳定）。RNA 发夹的后面通常跟有一串 4 个或更多的 U 残基。聚合酶在合成发夹 RNA 后似乎就立刻停了下来。其后的一串 U 残基与反义 DNA 链的相应的 A 残基配对作用弱，这有利于 RNA 与模板 DNA 链的解离，结果 RNA 从转录复合物中释放出来。反义 DNA 链上非配对的部分随即与 DNA 有义链退火，核心酶从 DNA 上解离。

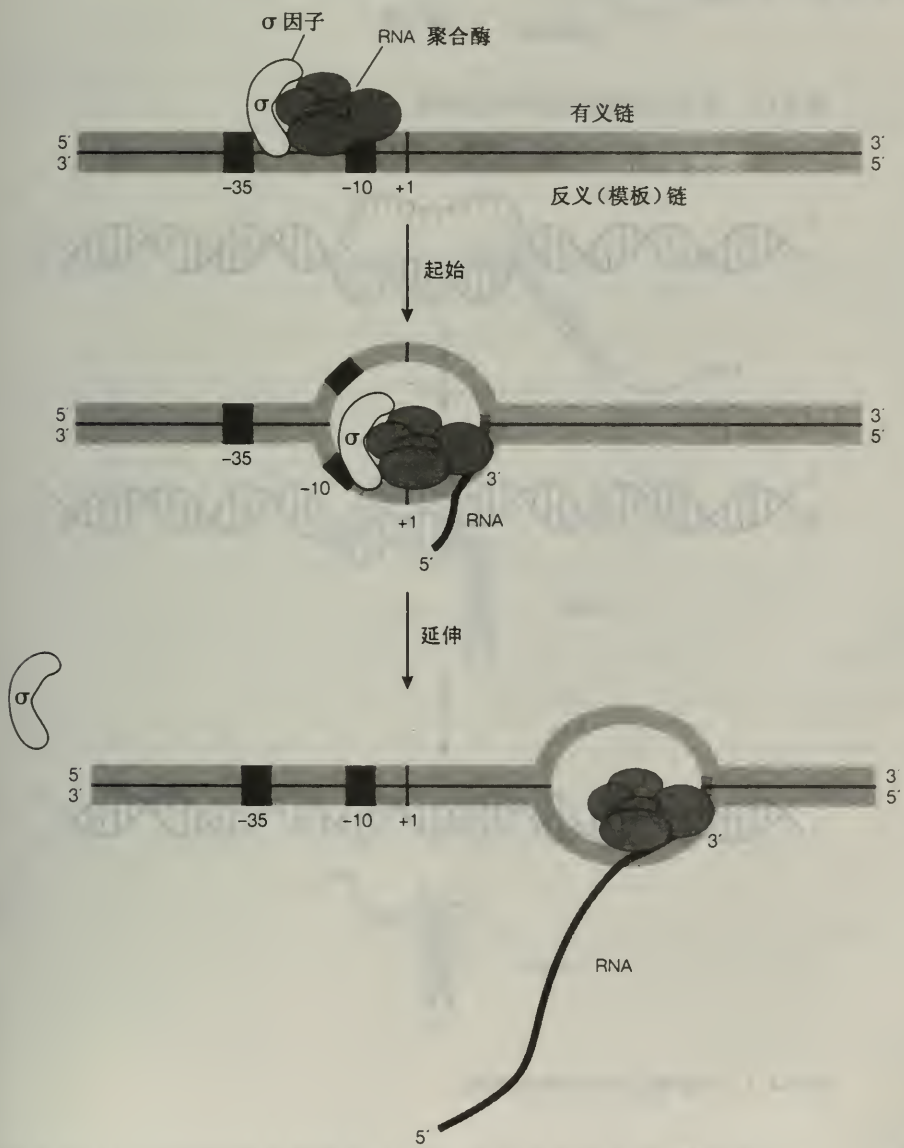


图 K4.1 转录复合体的形成：起始与延伸

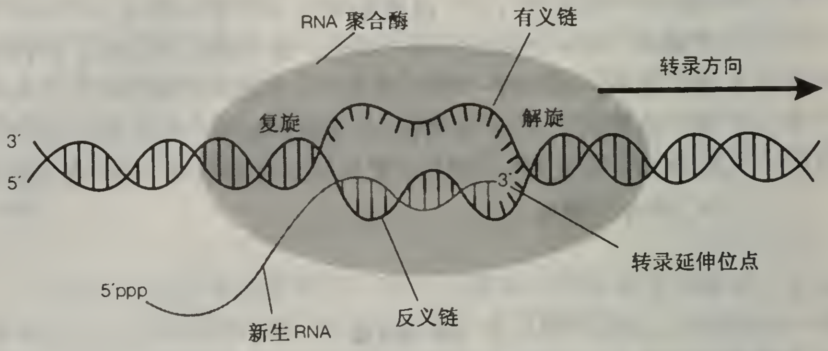


图 K4.2 延伸过程中转录泡的结构简图

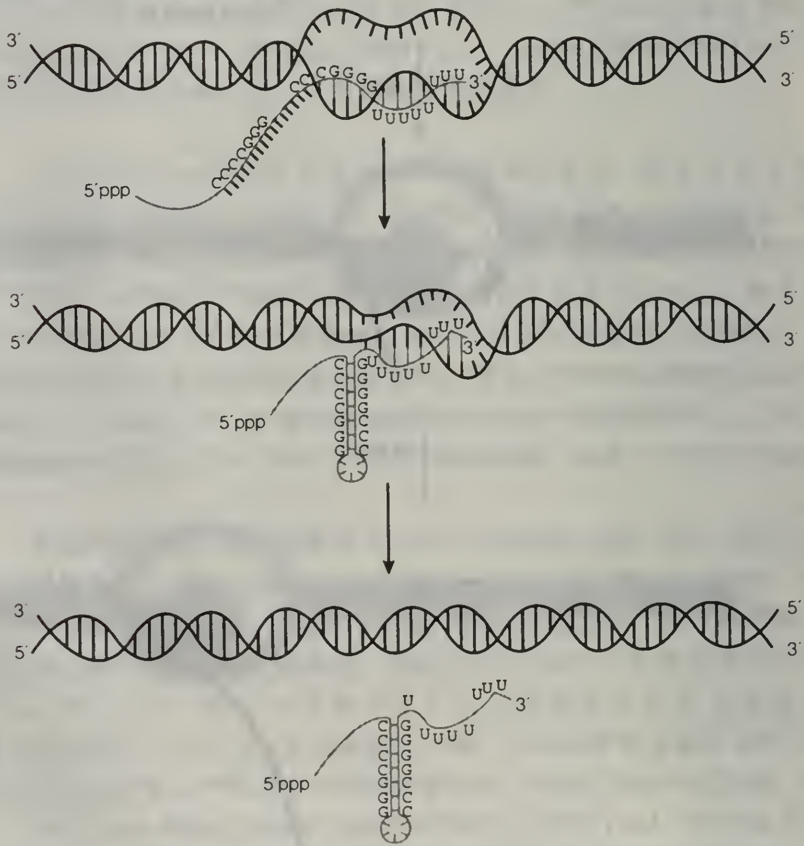


图 K4.3 不依赖 ρ 的转录终止简图

依赖 ρ 的转录终止

尽管 RNA 聚合酶可在紧跟一串 U 残基的发夹结构处自行终止转录，但有的终止位点并不形成强的发夹结构，须用一种辅助因子即 **ρ 蛋白**来帮助转录终止。ρ 是一个六聚体蛋白质，在有单链 RNA 存在时可水解 ATP。

这个蛋白似乎能结合 RNA 上一串 72 个核苷酸，这可能通过识别一种特定的结构特征，而不是某个共有序列。ρ 沿新生 RNA 链向转录复合物移动，能使 RNA 聚合酶在 ρ 依赖性终止子处终止转录。像不依赖 ρ 的终止子一样，这些信号是在新合成的 RNA 链中而不是在模板 DNA 链中被识别的。有时 ρ 依赖性的终止子有发夹结构但缺少之后的不依赖 ρ 终止所必需的一串 U 残基。

（程　凌　译　李文君　校）



L1 乳糖操纵子

要 点	
操纵子	1961 年 Jacob 和 Monod 最先提出了操纵子的概念，操纵子是原核生物基因表达的单元，它包括被协同调节的基因（结构基因）和被调节基因的产物所识别的调控元件。
乳糖操纵子	*lacZ*、*lacY* 和 *lacA* 基因均在单一个启动子 P_{lac} 的调控下从 *lacZYA* 转录单元被转录，编码将乳糖作为碳源所需的酶。*lacI* 基因的产物——乳糖阻抑物是 P_{lac}上游的另一个独立的转录单元表达形成的。
乳糖阻抑物	乳糖阻抑物由四个相同的蛋白亚基所组成。其上有一对称的结构，并能与 *lacZYA* RNA 起始位点重叠的一段 28 个碱基的具回文（对称）结构的操纵序列 O_{lac}相结合从而阻止从 P_{lac}的转录。
诱 导	乳糖阻抑物与诱导物（它的存在依赖于乳糖）结合后，其构象发生改变以致于无法再与 O_{lac} 操纵序列相结合，从而迅速诱导 *lacZYA* 的转录。
cAMP 受体蛋白	cAMP 受体蛋白（CRP）是一个转录激活因子，通过与 cAMP 的结合而被激活。葡萄糖缺乏时 cAMP 的水平升高，这个复合物结合到 P_{lac}上游的一个位点从而引起 DNA 链产生 90°的弯曲，使得 RNA 聚合酶结合到启动子上并起始转录。CRP 激活因子参与对葡萄糖水平应答的分解代谢操纵子的基因表达的全局调节。
相关主题	转录的基本原则（K1） 转录的起始、延伸与终止（K4） 大肠杆菌 RNA 聚合酶（K2） 色氨酸操纵子（L2） 大肠杆菌 σ^{70}启动子（K3）

操纵子

1961 年，Jacob 和 Monod 提出了操纵子模型，这是与特殊代谢途径有关的基因转录的协同调控模型。**操纵子**是基因表达和调控的单元，典型的操纵子包括：

- **结构基因**（除调节基因以外的所有基因），编码那些在某一特定的生物合成途径中起作用的、其表达被协同调控的酶。
- 调控元件，如**操纵序列**，是调节结构基因转录的一段 DNA 序列。
- 调节基因，其产物能够识别调控元件，例如阻抑物，可以结合并调控操纵基因序列。

乳糖操纵子

大肠杆菌能利用乳糖作为碳源，而利用乳糖作为碳源的酶只有当乳糖成为惟一的碳源时才会被合成。**乳糖操纵子**（或 *lac* 操纵子，图 L1.1）由三个结构基因组成：*lacZ*，编码 β-半乳糖苷酶，它可以将乳糖水解为半乳糖和葡萄糖；***lacY***，编码半乳糖苷透性酶，它能将乳糖运送透过细菌的细胞壁；***lacA***，编码硫代半乳糖苷乙酰转移酶。这三个结构基因被同一个转录单元 *lacZYA* 所编码，它有一个启动子 P_{lac}。这种结构意味着乳糖操纵子的这三个结构蛋白是作为一个**多顺反子的 mRNA**（在同一调控序列下含有不止一个编码区）一起表达的。*lacZYA* 转录单元含有一个**操纵基因位点 O_{lac}**，它位于 P_{lac} 启动子 5′端的 − 5 至 + 21 之间。该位点可被一个称作**乳糖阻抑物**的蛋白相结合。当该蛋白结合到操纵基因上时，便成为转录的强抑制物。乳糖阻抑物被一个独立的调节基因 ***lacI*** 所编码，*lacI* 也是乳糖操纵子的一部分；*lacI* 位于 P_{lac} 的上游。

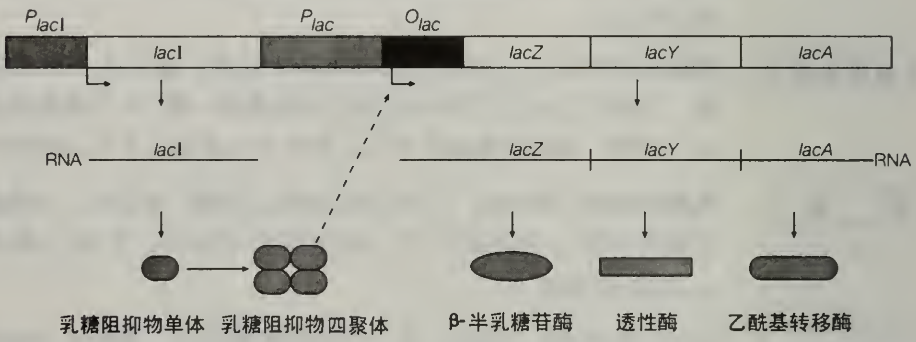


图 L1.1　乳糖操纵子的结构

乳糖阻抑物

lacI 基因编码乳糖阻抑物，当以同亚基四聚体形式存在时具有活性。它对于 *lac* 操纵序列结合位点 O_{lac} 具有很强的亲和力，并且对 DNA 也有较高的亲和力。*lac* 操纵序列位点由具有**回文结构**的 28 bp 碱基组成（回文结构是指当一条链以 5′到 3′的顺序从左向右读时其 DNA 序列与它的互补链也以 5′至 3′的方向从右向左读时完全相同参见 G3）。操纵序列的这种反向重复对称正好与由四个相同亚基组成的乳糖阻抑物的内部对称相匹配。当乳糖缺乏时，阻抑物占据了操纵序列的结合位点。乳糖阻抑物和 RNA 聚合酶似乎能够同时结合到乳糖启动子和操纵序列位点上。乳糖阻抑物实际上使聚合酶结合到乳糖启动子上的能力增加了两个数量级。这就意味着，当乳糖阻抑物结合到 O_{lac} 操纵序列上时，聚合酶也很可能结合到相邻的 P_{lac} 启动子序列上。

诱导

当缺少诱导物时，乳糖阻抑物阻碍了几乎全部的 *lac ZYA* 的转录而使之保持很低的转录水平。将乳糖加入细胞中后，细胞本身所含的低量的透性酶使它能吸收乳糖，β-半乳糖苷酶则催化一些乳糖转化为异乳糖（图 L1.2）。

异乳糖可以作为诱导物结合到乳糖阻抑物上。从而引起阻抑物四聚体构象的变化，从而降低其对乳糖操纵序列的亲和力（图 L1.3），乳糖阻抑物从操纵序列上脱离下来可以使聚合酶（已经位于邻近的启动子上）迅速开始 *lacZYA* 基因的转录。因此，加入乳糖或合成诱导物如**异丙基-β-D-硫代半乳糖苷**（IPTG）（图 L1.2）能非常迅速地刺激乳糖操纵子结构基因的转录。若随后将诱导物清除掉则会导致诱导性转录的立即终止，这是因为游离的乳糖阻抑物会立即重新占据操纵序列上的位点，而 *lac ZYA* RNA 的转录物是极不稳定的。

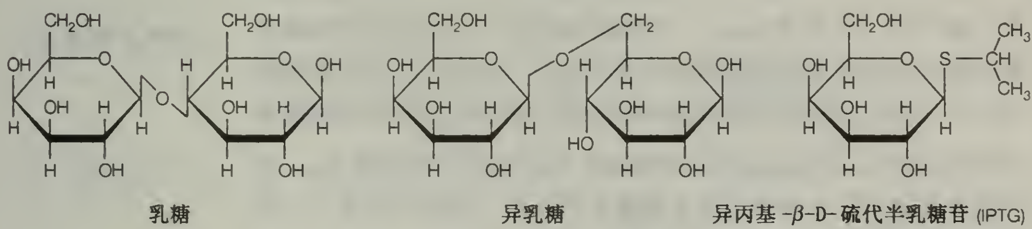


图 L1.2 乳糖、异乳糖和 IPTG 的结构

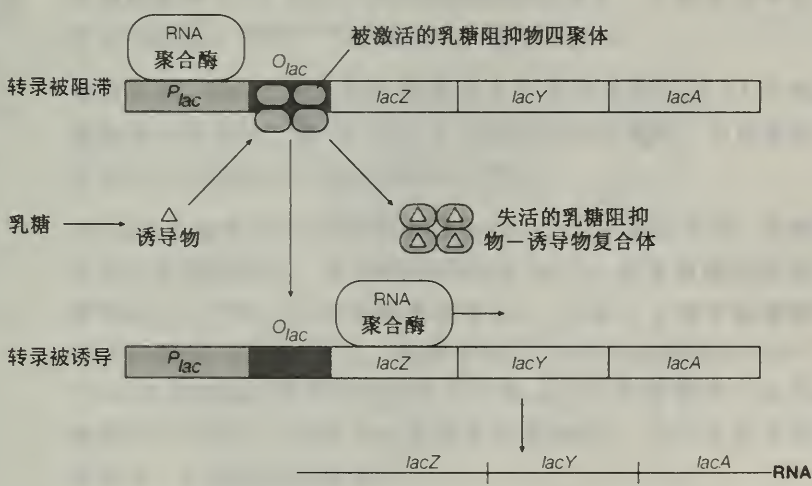


图 L1.3 诱导物的结合使乳糖阻抑物失活

cAMP 受体蛋白

P_{lac}启动子并非强启动子。P_{lac}及其相关的启动子没有强的 -35 序列，其中一些甚至只有弱的 -10 的共有序列。为了实现高水平的转录，它们需要一种特定的激活蛋白即 **cAMP 受体蛋白**（CRP）的活化。CRP 也可称为**分解代谢激活蛋白**或 CAP。当葡萄糖存在时，大肠杆菌不需要像乳糖这样的替代碳源。因此代谢操纵子如乳糖操纵子通常不被激活。这种调节由 CRP 介导，以二聚体形式存在的 CRP 自身不能独立与 DNA 结合，也不能调节转录。葡萄糖会降低细胞内 cAMP 的水平。当葡萄糖缺乏时，

大肠杆菌细胞内 cAMP 水平上升，CRP 结合到 cAMP 上。CRP-cAMP 复合物可结合到紧邻 RNA 聚合酶结合位点上游的乳糖操纵子的启动子 P_{lac} 上。CRP 的结合引起 DNA 链发生 90°的弯曲，这增强了 RNA 与启动子的结合，使转录效率提高 50 倍。

CRP 的结合部位是反向重复序列，可能位于启动子附近（就像在乳糖操纵子中那样）、也可能位于启动子内部、还可能位于远离启动子的上游。不同代谢操纵子的启动子上的 CRP 结合位点的差异可能导致这些操纵子在体内对 cAMP 作出不同程度的反应。

L2 色氨酸操纵子

要 点

色氨酸操纵子 色氨酸操纵子编码色氨酸生物合成途径中的5个结构基因。编码这5个酶的单一转录物是在单一启动子（P_{trp}）和操纵基因（O_{trp}）位点的调控下合成的。

色氨酸阻抑物 色氨酸阻抑物是一个单独的操纵子（*trpR* 操纵子）的产物。该阻抑物是一个二聚体，只有当它与色氨酸形成复合物后才能与色氨酸操纵基因相互作用。阻抑物的结合使转录效率下降了70倍。

弱化子 在 *trpE* 编码序列起始位点上游前面的162 bp的 *trp* 前导序列内有一个终止子序列，是一个不依赖 ρ 因子的终止子，在其发夹结构后跟着8个连续的U，能在+140 bp处终止转录。这一结构称为弱化子，因为它能导致色氨酸RNA合成的提前终止。

前导RNA结构 色氨酸前导RNA由四个序列互补的区域组成，可形成各种不同的发夹结构。其中一个就是弱化子的发夹结构。

前导肽 前导RNA有一个有效的核糖体结合位点并编码一段14个氨基酸残基的前导肽。第10与11号密码子编码色氨酸。当色氨酸含量低时，核糖体会在这些密码子上停顿。

弱化作用 RNA聚合酶在DNA模板的前导肽编码序列末端处停顿。核糖体起始前导肽翻译时，聚合酶继续转录RNA。如果核糖体在色氨酸密码子上停顿（比如色氨酸含量低），它减少了用于碱基配对的互补前导序列的供应，形成替代弱化子发夹结构的另一个DNA发夹结构。最终导致转录并不终止。如果核糖体不在色氨酸密码子上停滞（例如当色氨酸含量很高时），弱化子发夹就能够形成，转录将提前终止。

弱化的重要性 弱化作用使色氨酸的转录调控能提高10倍。转录弱化作用发生在至少6个与氨基酸生物合成相关的操纵子中。在一些操纵子（如组氨酸）中，这是氨基酸合成的反馈调节的惟一机制。

相关主题

转录的基本原则（K1）
转录的起始、延伸与终止（K4）
大肠杆菌RNA聚合酶（K2）
乳糖操纵子（L1）
大肠杆菌 σ^{70} 启动子（K3）

色氨酸操纵子

色氨酸操纵子包括色氨酸合成途经中相关的 5 个结构基因（图 L2.1）。该操纵子编码一个转录单元，即从色氨酸启动子 P_{trp} 和操纵基因位点 O_{trp} 向下游转录出 7 kb 的转录物。像许多涉及氨基酸生物合成的操纵子一样，色氨酸操纵子也具有当细胞内缺乏生物合成途径的产物色氨酸时，可使这些基因协同表达的进化系统。至于乳糖操纵子，其转录单元的 RNA 产物非常不稳定，使细菌能够对色氨酸的需求变化做出快速反应。

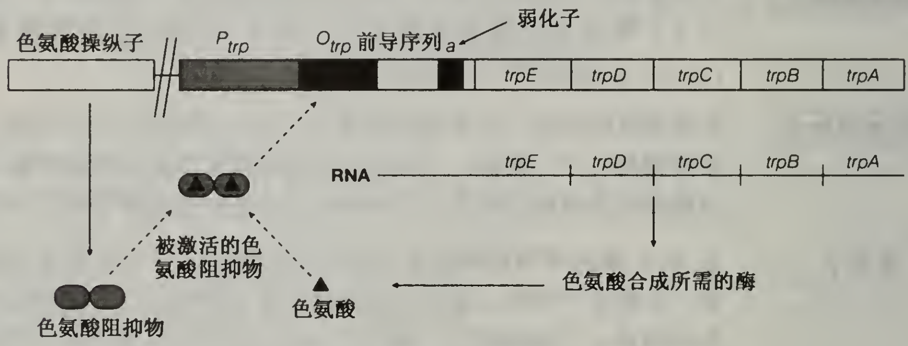


图 L2.1 色氨酸操纵子的结构和色氨酸阻抑物的功能

色氨酸阻抑物

独立的操纵子 *trpR* 编码的产物——色氨酸阻抑物，能与色氨酸操纵子的操纵基因位点特异性相互作用。这一形成色氨酸阻抑物结合位点的对称的操纵基因与色氨酸启动子序列在 −21 和 +3 碱基之间重叠。中心结合区是 18 个碱基的回文结构。色氨酸阻抑物能结合色氨酸并且只有当它与色氨酸结合形成复合物后才能结合到操纵基因上。阻抑物是含有两个亚基的二聚体，它与 CRP 蛋白和乳糖阻抑物具有相似的结构（参见 L1）。这一阻抑物二聚体含有一个中心区域和两个灵活的 DNA 解读头部（DNA-reading head），这两个解读头部分别由一个亚基的羧基端形成。只有当色氨酸结合到阻抑物上时，这些解读头部才会处于正确的距离，侧链才会处于正确的构象从而与色氨酸操纵基因的 DNA 连续的大沟相互作用（参见 C1），这样色氨酸操纵子所编码的酶的终产物，即色氨酸才可作为共阻抑物起作用，并且通过**终产物抑制**自身的合成。该阻抑物将转录的起始减少了大约 70 倍。这种对转录的影响作用比乳糖阻抑物结合介导的影响小得多。

弱化子

起初人们认为所有的色氨酸操纵子转录水平的调控都与阻抑物有关。然而据观察在操纵基因和 *trpE* 基因编码序列之间一段序列的缺失将导致基本转录水平和活化（解阻抑）后的转录水平的上升。该位点称为**弱化子**，它位于 *trpE* 起始密码子前面 162 nt 的转录前导序列的末端。弱化子是一个不依赖 ρ 因子的终止子区域，在一小段富含 GC 的回文结构之后是 8 个连续的 U 残基（参见 K4）。如果这段序列能在 RNA 转录物中形成发

夹结构，就可以作为一个高效的转录终止子，因而只有 140 bp 的转录物被合成。

前导 RNA 结构

色氨酸操纵子 RNA 的前导序列含有 4 个序列互补区，能形成不同的碱基配对的 RNA 结构（图 L2.2）。它们被称为序列 1、2、3 和 4。弱化子发夹结构是序列 3 和 4 配对的产物（3∶4 结构）。序列 1 和 2 也是互补的，能形成另一个发夹结构 1∶2。然而序列 2 还和序列 3 互补。如果序列 2 和 3 形成 2∶3 发夹结构，3∶4 弱化子发夹结构就不能形成，转录就不会终止。在正常情况下，1∶2 和 3∶4 发夹结构的形成在能量上是有利的（图 L2.2a）。

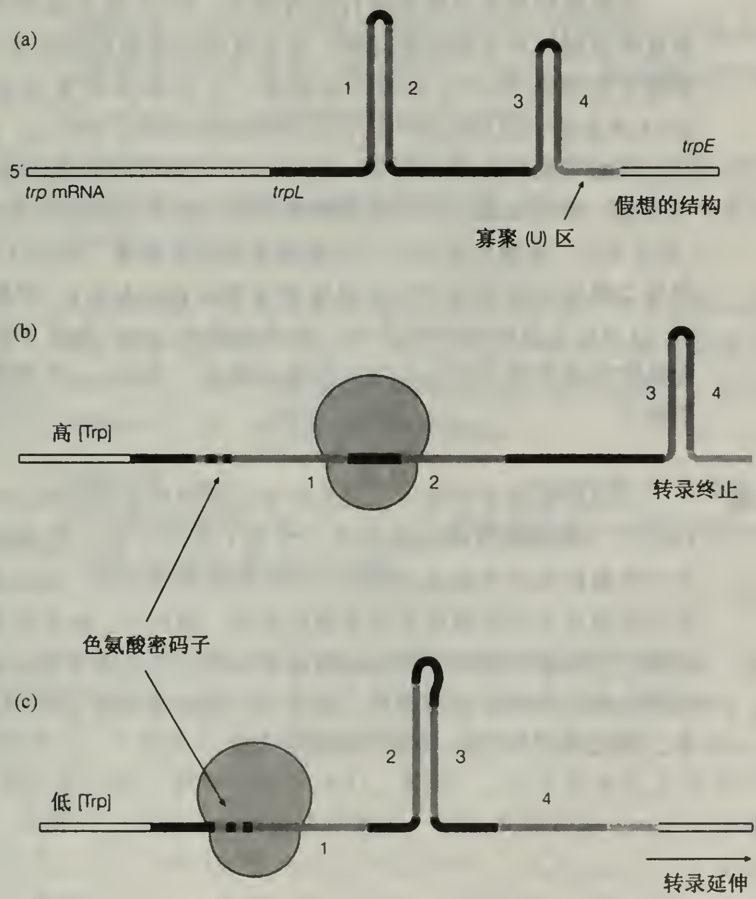


图 L2.2 色氨酸操纵子的转录弱化作用

前导肽

前导 RNA 序列中含有一个有效的核糖体结合位点并能形成由前导 RNA 27～68 号碱基所编码的 14 个氨基酸的前导肽。这段前导肽的第 10 和第 11 个密码子编码连续的色氨酸残基，即色氨酸操纵子合成酶的终产物。这条前导肽作为多肽链没有明显的功能，而且色氨酸是稀有氨基酸；

因此，两个色氨酸密码子连续出现的可能性很小，在色氨酸含量很低的情况下核糖体将会在这个位点停滞。前导肽的作用就是决定色氨酸的含量并控制转录的终止。

弱化作用

大肠杆菌中，弱化作用依赖于转录和翻译的紧密偶联，使翻译可以在mRNA边转录时边进行。色氨酸前导肽编码序列的3′端与互补序列1重叠（图L2.2），两个色氨酸密码子都在序列1内，终止密码子在序列1和2之间。由于色氨酸（色氨酸操纵子合成的酶的最终产物）是翻译过程中所必需的，因此细胞对它的含量很敏感，它决定了在mRNA中终止子(3∶4)发夹结构是否形成。

在色氨酸操纵子转录进行的过程中，RNA聚合酶在序列2的末端停滞直到核糖体开始翻译前导肽。在色氨酸含量很高的情况下，核糖体迅速在两个色氨酸密码子处插入色氨酸，这样翻译可直达前导肽末端。当RNA聚合酶到达终止序列时，于是核糖体封闭了序列2，3∶4发夹得以形成，转录可能被终止（图L2.2b）。这一过程被称为**弱化作用**。

另一种情况是：如果色氨酸缺乏时，翻译过程中将缺少其氨酰tRNA（参见P2），核糖体将在两个色氨酸密码子上滞留，封闭了序列1。这使得序列2能自由地与序列3形成发夹结构（图L2.2c），即**抗终止子**。终止子（3∶4）发夹结构不能形成，转录继续到*trpE*及其下游。因此终产物色氨酸含量的高低决定了转录是提早终止（弱化），还是继续转录完整个操纵子。

弱化的重要性

仅仅依靠弱化作用，色氨酸的存在就使色氨酸操纵子的转录被抑制了10倍。与色氨酸阻抑物的作用（70倍）合在一起，这就意味着色氨酸水平对色氨酸操纵子的表达施加了700倍的调节效果。至少在六种与氨基酸的生物合成相关的操纵子中有弱化作用。例如*his*操纵子含有编码一个有连续7个组氨酸密码子的多肽的前导序列。并不是所有这些操纵子都像色氨酸操纵子那样有协同调控。例如*his*操纵子就没有阻抑物-操纵基因调控，弱化作用是其惟一的反馈调控机制。

L3 不同 σ 因子对转录的调节

要 点

σ 因子	σ 因子负责识别启动子共有序列并且只为转录起始所需，许多细菌产生不同系列的 σ 因子。
启动子识别	在 *E.coli* 中，σ^{70} 负责识别 −10 和 −35 共有序列。不同的共有序列能够在被不同的 σ 因子调节的基因中找到。
热休克	大肠杆菌大约有 17 种蛋白在温度超过 37℃ 时可以特异地表达。这些蛋白通过利用另一种 σ 因子 σ^{32} 的 RNA 聚合酶的转录被表达。σ^{32} 有它自己特异的启动子共有序列。
枯草杆菌的孢子形成	在不适宜的环境条件下，枯草杆菌通过基本的细胞分化过程，即细胞分化成母细胞和前芽孢来形成芽孢。这一过程被一系列 σ 因子严格地调控。
噬菌体 σ 因子	很多噬菌体为了取代宿主细胞的转录而合成它们自身的 σ 因子代替正常宿主细胞的 σ 因子，从而改变 RNA 聚合酶对启动子序列的专一性。枯草杆菌 SPO1 噬菌体表达一系列 σ 因子，从而实现噬菌体早、中、晚期基因的顺序表达。

相关主题

细胞分类（A1）
大肠杆菌 σ^{70} 启动子（K3）
转录的基本原则（K1）
转录的起始、延伸与终止（K4）
大肠杆菌 RNA 聚合酶（K2）

σ 因子

RNA 聚合酶的核心酶 αββ′ω 不能在启动子位点开始转录（参见 K）。为了特异地识别一般启动子的共有序列 −35 和 −10，就需要有 σ 因子亚基的参与。该亚基只在转录起始中需要，当转录起始后进入 RNA 延伸阶段前从核心酶上释放（参见 K4）。因此，σ 因子好像是具双重作用的蛋白，既能结合到核心 RNA 聚合酶上，同时又能识别 DNA 上特殊的启动子序列。包括大肠杆菌在内的许多细菌产生一系列 σ 因子，它们能够识别不同类型的启动子。从单一启动子或一小组启动子开始的转录通常被单一的转录阻抑物（如乳糖阻抑物）或转录激活物（如 cAMP 受体蛋白，CRP）所调节。然而一些环境条件需要细胞内基因表达的总模式发生很大的改变。在这种条件下，细菌会利用另一组 σ 因子来指导 RNA 聚合酶结合到不同的启动子序列上。这一过程实现了细胞基本转录机制的分化以适应不同类型基因的特异性转录。

启动子识别

不同的 σ 因子与 RNA 聚合酶结合可以使细胞中负责一般 RNA 合成的聚合酶产生新的启动子特异性。通过比较聚合酶和专一的 σ 因子复合物活化的启动子，发现每个 σ 因子识别不同的大约以 − 35 和 − 10 位点为中心的序列组成。似乎 σ 因子本身与这两个区域都有联系，但 − 10 区域是最重要的。σ^{70}亚基是在大肠杆菌中最普遍的 σ 因子，它负责识别含有共有元件 − 35 和 − 10 的一般启动子。

热休克

对热休克的反应是大肠杆菌中基因表达通过使用不同的 σ 因子而发生很大变化的实例。如大肠杆菌被升温时，大约 17 个左右的蛋白，称作热休克蛋白被诱导合成。如果将大肠杆菌从 37℃ 转移到 42℃，热休克蛋白会在瞬间迅速地合成。然而，如果温度的上升更加剧烈，例如升到 50℃，大肠杆菌不能生长，热休克蛋白就成了惟一被合成的蛋白。大肠杆菌热休克蛋白编码基因的启动子被一个含有变异的 σ 因子（σ^{32}）的 RNA 聚合酶全酶所识别，σ^{32}由 *rpoH* 基因编码。σ^{32}是一个次要蛋白，其丰度比 σ^{70}小得多。含 σ^{32}的全酶主要作用于热休克蛋白的启动子但不识别大部分其他基因的共有序列启动子（图 L3.1）。相应地，热休克启动子含有一些与结合到 σ^{70}上的其他一般启动子的不同的序列。

保守启动子		-35 序列		-10 序列
标准 (σ^{70})	- - - - - - - - -	TTGACA	·······16–18 bp·······	TATAAT
热休克 (σ^{32})	T - - C - C - C	TTGAA	·····13–15 bp·····CCC	CAT-T

图 L3.1　热休克（σ^{32}）应答启动子和一般（σ^{70}）应答启动子的比较

枯草杆菌的孢子形成

营养生长的枯草杆菌细胞对不良环境作出反应时形成孢子（参见 A1）。孢子的形成（或孢子化）需要基因表达发生巨大改变，包括几乎所有营养生长所需要的蛋白合成均停止和在合适条件下孢子萌发所需的蛋白的合成。孢子形成的过程包括细菌细胞非对称地裂解为两部分，一部分为前孢子，它最终形成孢子，另一部分为母细胞，它最后被遗弃。这一体系被看作是细胞分化最基本的例子。枯草杆菌 RNA 聚合酶在功能上与大肠杆菌的一致。营养生长的枯草杆菌含有多种 σ 因子。孢子形成被另一套 σ 因子和营养生长细胞中的 σ 因子共同调控。在前孢子和母细胞中，细胞分化开始前不同的 σ 因子具有各自特殊的活性，这种区域化的交叉调节使芽孢和母细胞可以紧密地共同协调分化过程。

噬菌体 σ 因子

一些噬菌体能够产生新的 σ 亚基以提供给宿主 RNA 聚合酶不同的启动子序列特异性，以此来选择性地表达它们自身的噬菌体基因（例如大肠杆菌中的 T4 噬菌体和枯草杆菌中的 SPO1）。这一策略是噬菌体不需编码自身完整聚合酶的一种替代（例如 T7 噬菌体，参见 K2）。枯草杆菌中的

SPO1 噬菌体按**级联**表达一系列 σ 因子，使得在病毒感染的特定时期，它自身的基因可以被转录。开始时，早期基因被正常的细菌中的全酶所表达，在这些**早期基因**中含有编码 σ^{28}的基因，它随即代替了 RNA 聚合酶上细菌的 σ 因子。含有 σ^{28}的全酶负责**中期基因**的表达。噬菌体中期基因包括基因 *33* 和 *34*，它们能产生**后期基因**转录所需的 σ 因子。通过这种方法，噬菌体机械地运用宿主细胞的 RNA 聚合酶的机制按一定顺序表达它们自身的基因。

（郑 润 译 李文君 校）

M1 三种 RNA 聚合酶：性质与功能

要 点

真核生物 RNA 聚合酶

三种真核生物的聚合酶转录不同类型的基因，其活性可以由 3 种酶对真菌毒素——α-鹅膏蕈碱的不同敏感性来区分。

- RNA 聚合酶Ⅰ存在于核仁中，负责大多数前 rRNA 的合成。
- RNA 聚合酶Ⅱ存在于核质中，负责前 mRNA 和一些小核 RNA 的合成。
- RNA 聚合酶Ⅲ也存在于核质中，负责前 5S rRNA、tRNA、一些 snRNA 以及细胞质 rRNA 的合成。

RNA 聚合酶亚基

每种 RNA 聚合酶都含有 12 个或更多的不同亚基。最大的两个亚基彼此相似，并与 *E. coli* RNA 聚合酶的 β′和 β 亚基相似，其他亚基与 *E. coli* 聚合酶中的 α 亚基具有同源性。另外还有五种亚基在这三种聚合酶中是共同的，剩下的其他的亚基是各种聚合酶所特有的。

真核生物 RNA 聚合酶的活性

像原核生物 RNA 聚合酶一样，真核生物的聚合酶也不需要引物，沿 5′至 3′方向合成 RNA。与细菌聚合酶不同的是在与 DNA 结合时需要辅助因子。

RNA 聚合酶Ⅱ的 CTD

RNA 聚合酶Ⅱ的最大的亚基在羧基端有七个氨基酸的重复序列，被称为羧基末端结构域（carboxyl-terminal domain，CTD），序列 Tyr-Ser-Pro-Thr-Ser-Pro-Ser 在鼠的 RNA 聚合酶Ⅱ中重复 52 次，并且被磷酸化。

相关主题

蛋白质分析法（B3）
大肠杆菌 RNA 聚合酶（K2）
RNA 聚合酶Ⅰ基因：核糖体重复（M2）
RNA 聚合酶Ⅲ基因：5S 基因与 tRNA 基因的转录（M3）
RNA 聚合酶Ⅱ基因：启动子与增强子（M4）
通用转录因子与 RNA 聚合酶Ⅱ的起始（M5）
转录调控实例（N2）

真核生物 RNA 聚合酶

真核生物的转录机制与原核生物相似，然而与真核生物转录机制相关的大量多肽的参与使真核生物 RNA 聚合酶显得更复杂。三种不同的 RNA 聚合酶复合体负责不同类型真核生物基因的转录。不同的 RNA 聚合酶可

以依据在色谱层析柱上被洗脱下来的不同盐浓度而区分（参见 B4）。同时也可利用每一种 RNA 聚合酶对真菌毒素——α-鹅膏蕈碱有着不同的敏感性来区分它们的活性。

- RNA 聚合酶Ⅰ（RNA Pol Ⅰ）转录大部分 rRNA 的基因，存在于核仁中，并对 α-鹅膏蕈碱不敏感。
- RNA 聚合酶Ⅱ（RNA Pol Ⅱ）转录所有的编码基因和一些 snRNA 基因，存在于核质中，对 α-鹅膏蕈碱非常敏感。
- RNA 聚合酶Ⅲ（RNA Pol Ⅲ）转录 tRNA、5S rRNA 和 U6 snRNA 的基因以及其他一些小分子 RNA，也存在于核质中，对 α-鹅膏蕈碱中度敏感。

除了这些核中的酶外，真核生物的细胞还在线粒体和叶绿体中含有其他聚合酶。

RNA 聚合酶亚基

所有这三种聚合酶都是含有 12 个或更多亚基的大分子。编码每一 RNA 聚合酶的两个大亚基（相关的 DNA 编码序列）具有同源性。这三种真核生物聚合酶都含有与*E. coli* 核心 RNA 聚合酶 $\alpha_2\beta\beta'$亚基相同源的亚基（参见 K2）。每一真核生物 RNA 聚合酶中的最大亚基与*E. coli* 聚合酶中的 β′亚基相似，第二大亚基与*E. coli* 的含有活性位点的 β 亚基相似。这一活性位点同样存在于第二大亚基中，证明了这种同源性在功能上的重要性。在 RNA 聚合酶Ⅰ与聚合酶Ⅲ中共存两个亚基，另一个 RNA 聚合酶Ⅱ专有亚基与*E. coli* RNA 聚合酶的 α 亚基具有同源性。至少还有五种小亚基普遍存在于这三种不同的聚合酶中，此外仅存于某一特定聚合酶中的亚基还有 4～7 个。

真核生物 RNA 聚合酶的活性

与细菌 RNA 聚合酶一样，每种聚合酶都沿着 5′至 3′方向催化转录，并合成与反义模板链互补的 RNA。这一反应需要前体核苷酸 ATP、GTP、CTP 以及 UTP，不需要转录起始的引物，不同于纯的细菌聚合酶，纯的真核生物 RNA 聚合酶在与启动子结合、起始转录之前，需要一些额外的起始蛋白的参与。

RNA 聚合酶Ⅱ的 CTD

在 RNA 聚合酶Ⅱ的羧基端含有一段 7 氨基酸的序列。该序列在鼠的聚合酶中重复 52 次，在酵母中重复 26 次。这一七肽序列为 **Tyr-Ser-Pro-Thr-Ser-Pro-Ser**，被称作**羧基末端结构域**即 CTD，对酶活性是必需的。CTD 序列可在丝氨酸和某些络氨酸处磷酸化。体外研究表明在转录起始时，CTD 被去磷酸化，但在转录延伸过程中，随着 RNA 聚合酶解离启动子时，又被磷酸化。因为 RNA 聚合酶Ⅱ催化所有真核生物编码基因的转录，可见该酶对于研究特异基因的表达非常重要。现已发现 CTD 是转录延伸过程中特异激活的一个重要标靶（参见 M5 与 N2）。

M2 RNA 聚合酶Ⅰ基因：核糖体重复

要 点

核糖体 RNA 基因

前 rRNA 转录单元含有编码 18S、58S 和 28S rRNA 的三段序列，在基因组中成簇排列，中间被非转录的间隔序列分开。

核仁的作用

前 rRNA 由 RNA 聚合酶Ⅰ（RNA Pol Ⅰ）在核仁中合成。rRNA 基因排列成环一起组成核仁，也就是熟知的核组织区域。

RNA 聚合酶Ⅰ启动子

前 rRNA 启动子由两个转录控制区域组成。转录起始位点包含于核心元件上，上游调控元件（UCE）长约 50 个碱基，开始于 −100位点处。

上游结合因子

上游结合因子（UBF）与 UCE 结合，同时也可结合到核心元件上游的不同位点上，使得两位点间的 DNA 形成环状结构。

选择因子 1

选择因子 1（SL1）与 UBF-DNA 复合物结合并使其稳定，然后 RNA 聚合酶Ⅰ再结合进来，并起始转录。

TBP 与 TAF_I

SL1 由四个亚基组成，其中包括 TATA 结合蛋白（TBP）。TBP 为所有 3 种 RNA 聚合酶起始转录所需。其他一些因子是 RNA 聚合酶Ⅰ所特异的 TBP 相关因子，被称作 TAF_I。

其他 rRNA 基因

Acanthamoeba 含有一个简单的转录调控系统，包括一个单一的控制元件和一个单一因子 TIF-1，这些为 RNA 聚合酶Ⅰ的结合及在 rRNA 启动子上的起始所必需。

相关主题

基因组复杂度（D4）
转录的起始、延伸与终止（K4）
大肠杆菌 RNA 聚合酶（K2）
三种 RNA 聚合酶：性质与功能（M1）
大肠杆菌 σ^{70}启动子（K3）
rRNA 加工与核糖体（O1）

核糖体 RNA 基因

RNA 聚合酶Ⅰ负责 rRNA 在分裂间期的连续合成。人类细胞在不同染色体上分布有 5 簇 rRNA 重复基因，每簇大约有 40 个拷贝的 rRNA 基因（参见图 M2.1 与 D4）。每一 rRNA 基因产生一个 45S 的 rRNA 转录物，长约 13 000nt（参见 D4）。这一转录物随后被切成 28S RNA（5000nt）、18S（2000nt）和 5.8S（160nt）rRNA 各一个（参见 O1）。这种 RNA 多基因拷贝的连续转录是产生足够量且加工的 rRNA、进而包装进核糖体所必需。

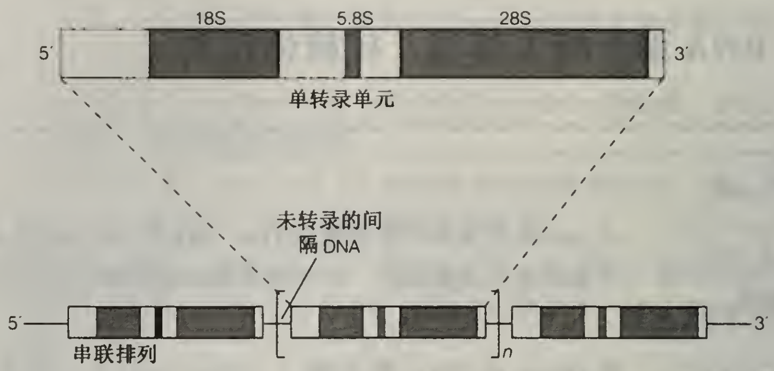


图 M2.1 核糖体 RNA 转录单元

核仁的作用

每一 rRNA 簇被称为一个核组织区域，因为核仁内含有与该基因簇相对应的 DNA 的大环。经有丝分裂形成一个细胞后，重新开始合成 rRNA，并在 rRNA 基因所在的染色体部位出现小核仁。在 rRNA 合成的活跃阶段，前 rRNA 转录物与 rRNA 基因捆扎在一起可在电子显微镜下看到**"圣诞树"样的结构**。在这些结构中，RNA 转录物与 DNA 被紧密地裹在一起，并从 DNA 上垂直突出。可以在基因起始处观察到短的转录物，它逐渐伸长直至转录单元末端，这可以从 RNA 转录物消失的现象中得到暗示。

RNA 聚合酶 I 启动子

哺乳动物的前 rRNA 基因启动子由两部分的转录控制区域构成（图 M2.2）。核心元件包括转录起始位点，跨越 −31 到 +6 区域。这一序列为转录所必需的。另一大约 50～80 个碱基区域称作**上游控制元件**（UCE），从起始上游大约 100bp 处（−100）起始。与核心元件单独作用相比，UCE 的存在使转录效率提高了 10～100 倍。

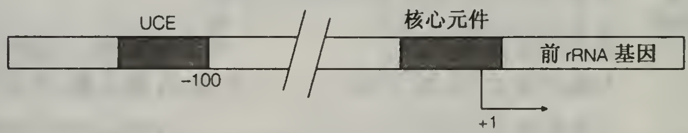


图 M2.2 哺乳动物前 rRNA 启动子的结构

上游结合因子

上游结合因子或 UBF 是一种特异 DNA 结合蛋白，可以与 UCE 结合。除了与 UCE 结合外，UBF 还可与核心元件上游的一段序列结合。两个 UBF 结合位点的序列间没有明显的相似性。一般认为 UBF 的一个分子结合一个序列元件。随后两个 UBF 分子可通过蛋白-蛋白相互作用而相互结合，导致在两个结合位点间的 DNA 形成一个环状结构（图 M2.3）。当 UBF 缺乏时可以观察到低水平的本底转录，而在 UBF 存在下，转录速率

会大大增强。

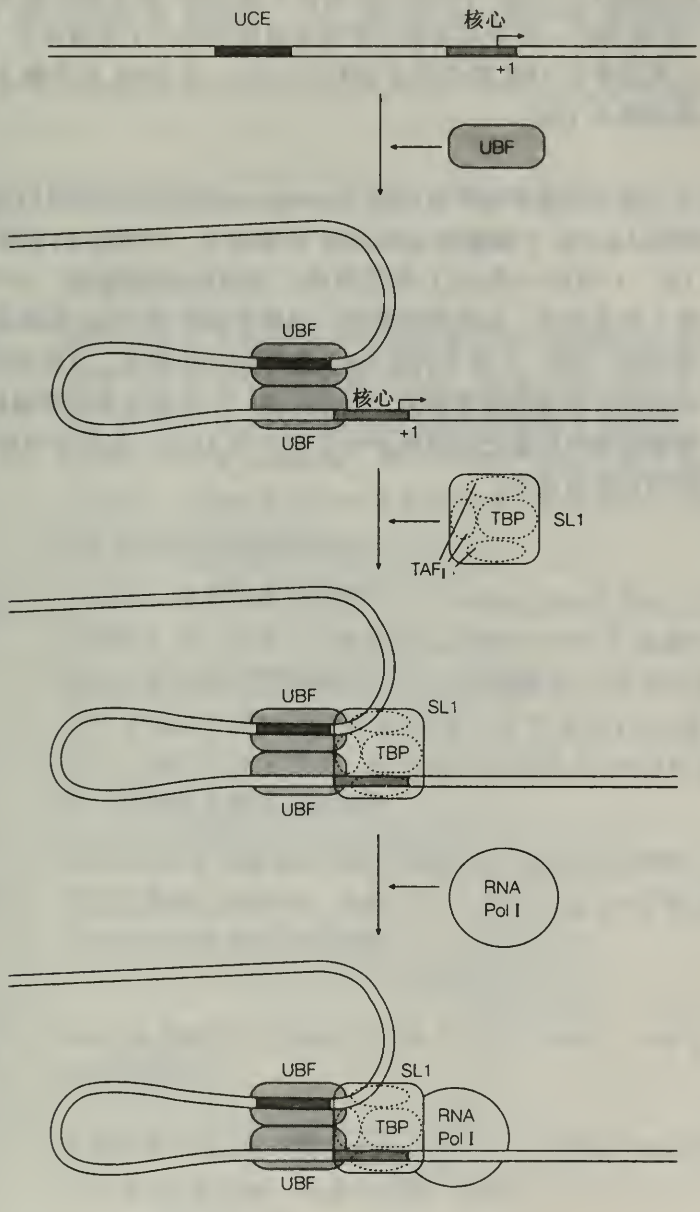


图 M2.3 rRNA 转录起始的图解模型

选择因子 1 **选择因子**（SL1）为 RNA 聚合酶 I 转录所必需。SL1 结合并稳定 UBF-DNA 复合物，且与核心元件游离的下游部分相互作用。SL1 的结合使得 RNA 聚合酶 I 能与上述复合物结合并起始转录，对 rRNA 转录非常重要。

TBP 和 TAF_I

研究发现 SL1 含有多个亚基，其中包括被称作 TBP 的蛋白（TATA 结合蛋白）。TBP 为所有三种真核生物 RNA 聚合酶起始所需（参见 M1、M3 和 M5），似乎是真核生物转录过程中的一个关键因子。SL1 的其他三个亚基属于 TBP 相关因子或称为 TAF，为 RNA 聚合酶Ⅰ转录所需的亚基被称为 TAF_I。

其他 rRNA 基因

在一个简单的真核生物 *Acanthamoeba* rRNA 基因的转录起始点上游大约 12～72 个碱基处的启动子区域内有一个单一的控制元件，被称作 TIF－1 的因子是 SL1 的同源物，与该位点结合后，方可让 RNA 聚合酶Ⅰ结合进来，从而起始转录。当聚合酶沿着 DNA 远离起始位点的方向移动时，TIF－1 因子仍与 DNA 处于结合的状态，这样会让另一聚合酶的起始进入转录的多轮循环。因此这是一个非常简单的转录控制系统。而在脊椎动物中好像已经进化产生了另一个 UBF，该因子负责将 SL1 的特异序列引向启动子。

M3 RNA 聚合酶Ⅲ基因：5S 基因与 tRNA 基因的转录

要 点

RNA 聚合酶Ⅲ	RNA 聚合酶Ⅲ（RNA Pol Ⅲ）由 16 个或更多的亚基组成，存在于核质中，负责转录 5S rRNA 的前体、tRNA 以及其他 snRNA 和胞质 RNA。
tRNA 基因	称作 A 框和 B 框的两个转录控制区位于转录起始点的下游，其序列不仅是 tRNA 中的保守序列，也是 DNA 启动子中的保守序列。TFⅢC 与 tRNA 启动子中的 A 框和 B 框结合；TFⅢB 与 TFⅢC-DNA复合体结合，并与 TFⅢC 结合位点上游的 DNA 相互作用。TFⅢB 含有三个亚基即 TBP、BRF 和 B″，并引导 RNA Pol Ⅲ的结合而起始转录。
5S rRNA 基因	5S rRNA 基因串联成簇排列，5S rRNA 启动子含有 2 个保守框即 C 框和 A 框。C 框位于起始位点下游 81～99 个碱基处，而 A 框则位于起始位点下游大约 50～65 个碱基处。TFⅢA 牢固地结合在 C 框启动子序列上，接着 TFⅢC 与 TFⅢA-DNA 复合体结合，并与 A 框序列相互作用。这一复合体形成后 TFⅢB 才能结合进来，吸收聚合酶并起始转录。
可变的 RNA 聚合酶Ⅲ启动子	许多 RNA 聚合酶Ⅲ启动子可被其上下游序列所调控，而其他启动子仅需要上游序列，包括 TATA 框以及在 RNA 聚合酶Ⅱ启动子中存在的其他序列来调控。
RNA 聚合酶Ⅲ的终止	RNA 聚合酶可在无辅助因子参与下终止转录，一串 A 残基常足以终止转录。
相关主题	大肠杆菌 RNA 聚合酶（K2） 三种 RNA 聚合酶：性质与功能（M1） RNA 聚合酶Ⅰ基因：核糖体重复（M2） RNA 聚合酶Ⅱ基因：启动子与增强子（M4） 通用转录因子与 RNA 聚合酶Ⅱ的起始（M5） rRNA 加工与核糖体（O1） tRNA 结构与功能（P2）

RNA 聚合酶Ⅲ

RNA 聚合酶Ⅲ（RNA Pol Ⅲ）是一个至少由 16 个不同亚基组成的复合体，与 RNA 聚合酶Ⅱ相同位于核质中。RNA 聚合酶Ⅲ负责 5S rRNA 的前体、tRNA 以及其他 snRNA 和胞质 RNA 的转录。

tRNA 基因　来自 tRNA 基因的最初转录物是被加工成成熟的前 RNA 分子（参见 O2）。tRNA 基因的转录控制区位于转录单元内的转录起始位点之后。在 DNA 编码的 tRNA 的 DNA 中有两个高度保守的序列，即 A 框（5′- TGGCNNAGTGG - 3′）和 B 框（5′- GGTTCGANNCC - 3′）。该序列同时也是一编码 tRNA 自身的重要序列，即编码 tRNA 的 D 环和 TΨC 环（参见 P2）。这意味着 tRNA 内的高度保守序列同时也是高度保守的启动子 DNA 序列。

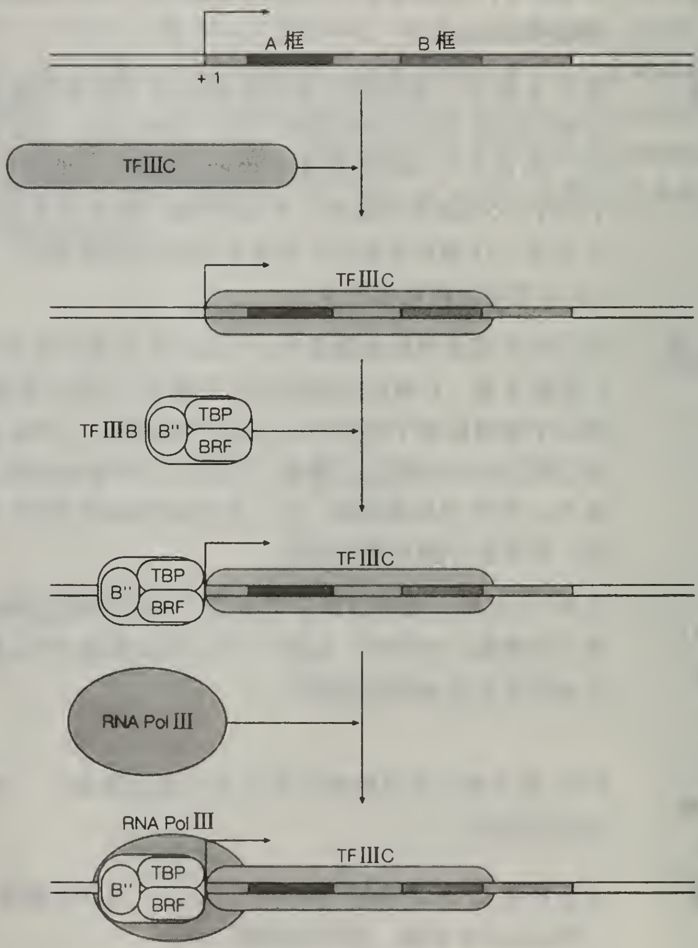


图 M3.1　真核生物 tRNA 启动子转录的起始

由 DNA 聚合酶Ⅲ起始 tRNA 基因转录所需的两个复杂 DNA 结合因子已被鉴定出来，即 **TFⅢB** 和 **TFⅢC**（图 M3.1）。TFⅢC 可与 tRNA 启动子的 A 框和 B 框结合，TFⅢB 则可与 A 框 50bp 上游序列结合。TFⅢB 由三个亚基组成，其中一个是 TBP，为三种 RNA 聚合酶所需（参见 M2 和 M5）。第二个亚基被称作 BRF（TFⅡB 相关因子，因为它与 RNA 聚合酶Ⅱ起始因子 TFⅡB 同源，参见 M5）。第三个亚基称作 B″。TFⅢB 没有

序列特异性，因此它的结合位点似乎由 TFⅢC 与 DNA 结合的位置所决定。TFⅢB 可使 RNA 聚合酶Ⅲ结合进来并起始转录。一旦 TFⅢB 结合后，TFⅢC 可在不影响转录的情况下解离出去。因此 TFⅢC 是一个为起始因子 TFⅢB 定位的装配因子。

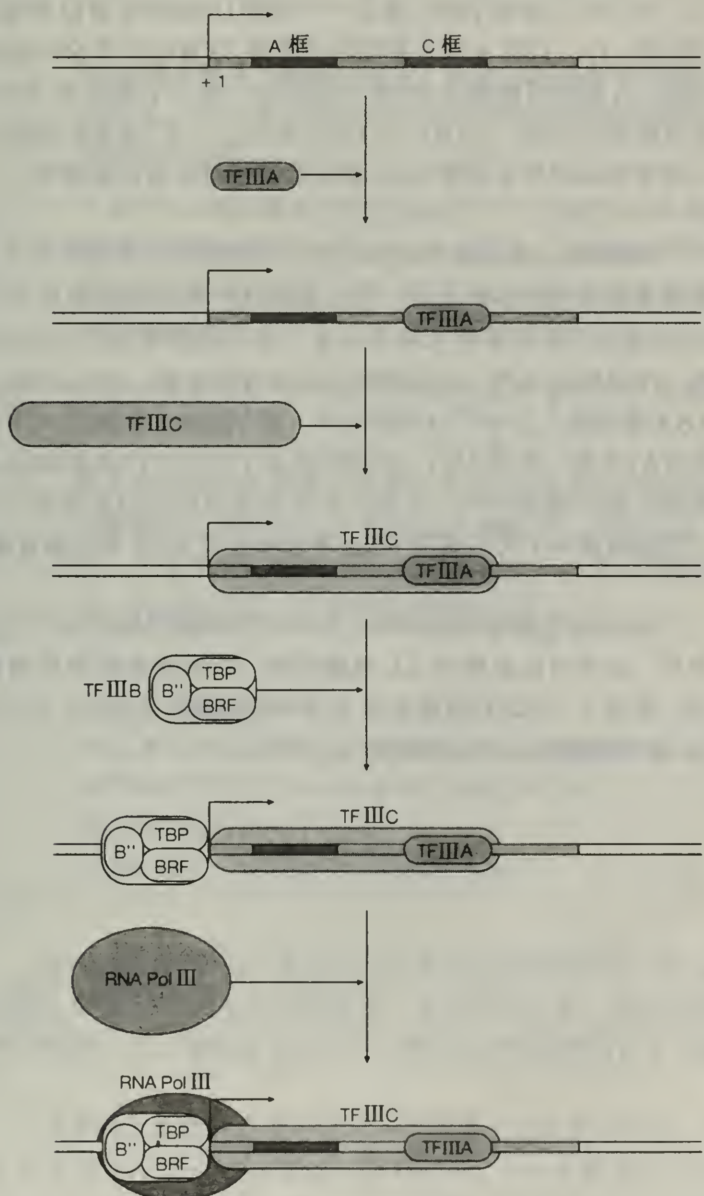


图 M3.2 真核生物 5S rRNA 启动子转录的起始

5S rRNA 基因

RNA 聚合酶Ⅲ转录核糖体大亚基的 5S rRNA 部分，是惟一单独被转录的 rRNA 亚基。与 RNA 聚合酶Ⅰ转录的另一些 rRNA 基因一样，5S

rRNA 基因也是串联排列在基因簇中。在人类基因组中，有一个大约由 2000 个基因组成的基因簇。5S rRNA 基因的启动子上含有一个称为 C 框的内部控制区域，它位于转录起始点下游 81～99 碱基处，而另一个位于 +50 至 +65 间的称为 A 框的序列也非常重要。

5S rRNA 启动子的 C 框是一个特异 DNA 结合蛋白 **TFⅢA** 的结合位点（图 M3.2）。TFⅢA 作为装配因子使得 TFⅢC 可与 5S rRNA 启动子相互作用。A 框也可能稳定 TFⅢC 的结合，在一个相对于 tRNA 启动子起始位点的等同位置，TFⅢC 与 DNA 结合。一旦 TFⅢC 结合，TFⅢB 就能与复合体起作用并促使 RNA 聚合酶Ⅲ的结合而起始转录。

可变 RNA 聚合酶Ⅲ启动子

许多 RNA 聚合酶Ⅲ基因也受上游序列作用来调控其自身转录。一些启动子诸如 U6 小核 RNA（U6 snRNA）和来自非洲淋巴瘤病毒中的小 RNA 基因只使用其转录起始位点上游的调控序列。U6 snRNA 的编码区有一个独特的 A 框，但该序列并不为转录所需。U6 snRNA 上游序列含有 RNA 聚合酶Ⅱ启动子的典型序列，其中包括一个位于 −30～−23 碱基处的 TATA 框（参见 M4）。在这些启动子中也与许多由 RNA 聚合酶Ⅱ所转录的U RNA基因一样，存在一些上游转录因子结合序列。这些现象表明一般的转录因子可以调节 RNA 聚合酶Ⅱ和 RNA 聚合酶Ⅲ基因。

RNA 聚合酶Ⅲ的终止

RNA 聚合酶Ⅲ转录的终止似乎仅需要聚合酶识别一个简单的核苷酸序列。该序列由成簇的 dA 残基构成，其终止效率受到周围的序列的影响。例如 5′-GCAAAAGC-3′序列就是在 *Xenopus borealis* 体细胞中 5S rRNA基因的一个有效的终止信号。

M4 RNA 聚合酶Ⅱ基因：启动子与增强子

要 点

RNA 聚合酶Ⅱ	RNA 聚合酶Ⅱ（RNA PolⅡ）催化所有编码基因的 mRNA 前体的合成。RNA 聚合酶Ⅱ转录的前 mRNA 须经过加帽、加尾和剪接等加工过程。
启动子	许多 RNA 聚合酶Ⅱ启动子都含有一个称做 TATA 框的序列，该序列位于起始位点上游 25～30 碱基处。其他的基因含有一个与起始位点重叠的起始元件，这些元件为基本转录复合体形成和转录起始所需。
上游调控元件	启动子上游 100～200bp 内的元件通常是有效转录所需的，例如 SP1 的 CCAAT 框。
增强子	能从启动子的上游或下游数千个 bp 处来激活转录的序列元件称为增强子。增强子可以是组织特异性的，也可以具有广泛的活性，并拥有多种基序。从极长的增强子元件到较短的启动子元件，其内均存在一系列连续的调控元件。
相关主题	大肠杆菌 σ^{70}启动子（K3） RNA 聚合酶Ⅰ基因：核糖体重复（M2） RNA 聚合酶Ⅲ基因：5S 基因与 tRNA 基因的转录（M3） 通用转录因子与 RNA 聚合酶Ⅱ的起始（M5） 真核生物的转录因子（N1） mRNA 加工、hnRNP 与 snRNP（O3）

RNA 聚合酶Ⅱ

RNA 聚合酶Ⅱ位于核质中，负责所有编码基因和一些 snRNA 基因的转录。前 mRNA 合成之后要经过一系列加工过程，包括在 RNA 5′端生成帽子结构、在 3′端加上 poly（A）尾巴以及剪接掉内含子（参见 O3）。

启动子

很多真核生物的启动子含有一个称做 **TATA 框**的序列，位于转录起始位点上游大约 25～35bp 位置（图 M4.1）。TATA 框有 7 对共有碱基序列：5′- TATA（A/T）A（A/T）- 3′，现在已经知道与 TATA 框结合的蛋白 TBP 结合 8 对碱基，包括一对额外的下游碱基，这一碱基对是什么并不重要（参见 M5）。TATA 框的作用与 *E. coli* 启动子的 - 10 序列相似，决定 RNA 聚合酶Ⅱ的位置或正确起始转录（参见 K3）。TATA 框与转录起始位点之间的序列并不重要，而 TATA 框周围的碱基序列却非常关键。

不过 TATA 框与转录起始位点之间的距离也很重要。转录起始位点的大约 50％是以嘌呤碱基开始。

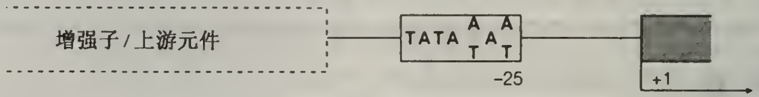


图 M4.1　含有 TATA 框的 RNA 聚合酶Ⅱ启动子

某些真核基因仅含有一个**起始元件**，而不是 TATA 框。该起始元件位于转录起始位点附近。许多起始元件在－1 位置是 C，在＋1 位置是 A。还有一些启动子既无 TATA 框又无起始元件，这些基因通常转录速率很低，其转录起始可发生在长至 200bp 内的不同起始点。这些基因通常在起始位点上游 100～200bp 范围内含有一段富含 20～50bp 的 GC 区。

上游调控元件

位于启动子上游的其他元件可以极大地提高基本启动子的低水平转录活性，这些元件在许多基因中存在，这些基因在不同的组织中表达水平相差很大。两个常见的例子是 **SP1 框**和 **CCAAT 框**，前者在很多基因的上游单独存在或与 TATA 框一道存在。启动子可以含有上述序列的一个、两个或多个拷贝。这些通常位于启动子上游 100～200bp 范围内的序列被称为**上游调控元件**（UREs），在确保启动子的有效转录方面起着重要作用。

增强子

许多真核生物启动子的转录可被远离转录起始位点数千个碱基的调控元件所增强，这一调控元件被称为增强子。这一现象最初是在 DNA 病毒 SV40 基因组中发现的，来自 SV40 DNA 的约 100bp 的序列，即使处于上游很远的位置也能显著增强基本启动子的转录。增强子的特征是 100～200bp 长，含有多个对增强子总体活性起作用的序列元件，这些元件或是广谱的，或是细胞类型特异性的。一般来说增强子具有下列共同特性：

- 赋予其相连基因从正确的起始位点的强转录活性；
- 不管位于其连接的基因的上游或下游任一方向上均可激活转录；
- 无论是上游还是下游，可在远离起始位点 1kb 以上发挥作用；
- 优先激活两个串连的启动子中距离最近的一个。

然而，随着更多的增强子和启动子被鉴定，现已发现上游启动子和增强子元件在结构和功能上均有重叠。在传统的增强子元件与那些具有方向特异性的启动子元件之间似乎存在一个连续体，必须放在靠近启动子的位置才有增强转录活性的作用。

M5 通用转录因子与 RNA 聚合酶Ⅱ的起始

要 点

RNA 聚合酶Ⅱ的基本转录因子	一系列复杂的基本转录因子，具有与 RNA 聚合酶Ⅱ启动子结合并一起起始转录的特征。而这些因子及其组成亚基正在被逐个鉴定，它们最初被命名为 TFⅡA、TFⅡB、TFⅡC 等。
TFⅡ D	TFⅡ D 与 TATA 框结合，是由 TATA 结合蛋白（TBP）与被称做 TBP 相关因子或 $TAF_{Ⅱ}$ 的多辅助因子构成一个多蛋白复合体。
TBP	TBP 是所有三种 RNA 聚合酶转录起始所需的转录因子。该因子有一个可与 TATA 框处的 DNA 小沟结合的马鞍型结构，可使 DNA 解旋并引入一个 45°的弯曲。
TFⅡ A	TFⅡA 可增强 TFⅡD 与 TATA 框的结合。TFⅡA 似乎可以中止抑制因子与 TFⅡD 的结合，否则这些抑制因子会进一步阻碍转录复合体的组装。
TFⅡB 与 RNA 酶的结合	TFⅡB 与 TFⅡD 结合，并为 RNA 聚合酶结合起一个桥梁作用。RNA 聚合酶与 TFⅡF 相连的复合体结合。
RNA 聚合酶进入之后的结合因子	RNA 聚合酶进入之后，TFⅡE、TFⅡH 和 TFⅡJ 按一定的结合顺序与转录复合体相结合。这些蛋白都是体外转录所必需的。
TFⅡH 作用下的 CTD 磷酸化	TFⅡH 使 RNA 聚合酶Ⅱ的羧基末端结构域（CTD）磷酸化，结果导致聚合酶复合体的形成。
起始转录复合体	TBP 在另一个 DNA 结合蛋白的作用下被添加到含起始子的启动子上，然后 TBP 就能以类似含 TATA 框的启动子机制起始转录。

相关主题

大肠杆菌 RNA 聚合酶（K2）
转录的起始、延伸与终止（K4）
三种 RNA 聚合酶：性质与功能（M1）
RNA 聚合酶Ⅰ基因：核糖体重复（M2）
RNA 聚合酶Ⅲ基因：5S 基因与 tRNA 基因的转录（M3）
RNA 聚合酶Ⅱ基因：启动子与增强子（M4）
真核生物的转录因子（N1）
转录调控举例（N2）

RNA 聚合酶Ⅱ的基本转录因子

一系列核转录因子已经被鉴定、纯化与克隆。这些多亚基因子均为RNA聚合酶Ⅱ启动子序列进行体外基本转录的起始所需，且被命名为转录因子ⅡA、ⅡB（TFⅡA、TFⅡB）等等。现已证明它们以特定的顺序在基本启动子上组装（图 M5.1），且可能受到不同水平的调控（参见 N1）。

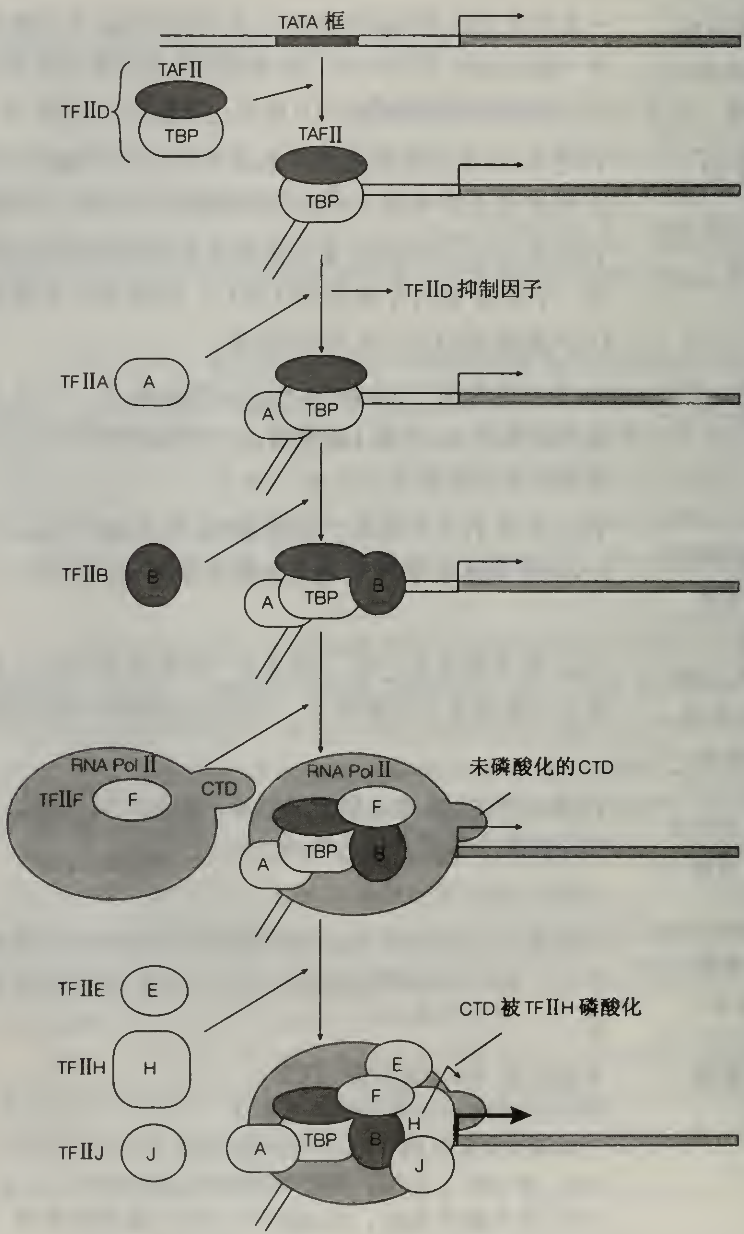


图 M5.1 RNA 聚合酶Ⅱ转录起始复合体在一含有 TATA 框的启动子上的组装

TFⅡD

在含有 TATA 框的启动子上，RNA 聚合酶Ⅱ转录因子 TFⅡD 负责与这个关键的启动子元件结合。TFⅡ D 与 TATA 框的结合是在 RNA 聚合酶Ⅱ转录起始复合体形成的早期。TFⅡD 是多蛋白复合体，复合体中只有一条多肽链即 **TATA 结合蛋白**（TBP）与 TATA 框结合，该复合体还含有其他多肽如 **TBP 相关因子**（TAF_{II}）。有迹象表明哺乳动物细胞中，是 TBP 与 TATA 框结合，接着至少 8 个 TAF_{II} 加入进来形成 TFⅡ D。

TBP

TBP 在真核生物三种转录复合体（SL1、TFⅢB 和 TFⅡD）中都存在，在转录起始中起重要作用（参见 M2 和 M3）。TBP 是一个单体蛋白，所有被分析过的真核生物的 TBP 都有一个由 180 个残基构成的非常保守的羧基末端结构域，这个保守的结构域在体内转录中的作用与全长蛋白所起作用相同。不太保守的羧基末端结构域的功能还不完全了解。TBP 有一个完全二元对称的马鞍型结构，但分子的两半却不完全相同。TBP 作用于 DNA 小沟，结果马鞍型结构的内部与 DNA 的 TATA 框结合，其蛋白的外表面便可与其他蛋白因子相互作用。TBP 的结合使 DNA 变形，导致 DNA 弯进马鞍型的内部并解链。这样就在 TATA 框 8 个碱基的头两个与最后两个碱基对之间形成一个 45°的扭结。虽然在 TATA 框结合的区域内发生突变的 TBP 具有为 RNA 聚合酶Ⅰ和Ⅲ转录所需的功能（参见 M2 和 M3），但它却抑制 RNA 聚合酶Ⅱ的转录起始。这表明另有两个聚合酶利用 TBP 起始转录，但是 TBP 在这些复合体中的确切作用仍不清楚。

TFⅡA

TFⅡA 与 TFⅡD 结合，并增强 TFⅡ D 与 TATA 框的结合，稳定 TFⅡ D-DNA 复合体。TFⅡA 至少由三个亚基组成。在体外转录研究中，在 TFⅡD 得到纯化后，就不再需求 TFⅡA 了。在完整的细胞中，TFⅡA 似乎能抵消抑制因子诸如与 TFⅡD 相关的 DR1 和 DR2 的作用，很可能 TFⅡA 与 TFⅡD 结合结果阻碍了这些抑制因子的结合，让组装过程得以继续。

TFⅡB 与 RNA 酶的结合

一旦 TFⅡD 结合到 DNA 上，另一个转录因子 **TFⅡB** 就会与 TFⅡD 结合，而 TFⅡB 又可与 RNA 聚合酶结合。这似乎是转录起始中重要的一步，因为 TFⅡ B 起一个允许聚合酶和另一个因子 **TFⅡF** 加入到复合体中来的桥梁作用。

RNA 聚合酶进入之后的结合因子

RNA 聚合酶结合之后，另三个转录因子，**TFⅡE**、**TFⅡH** 和 **TFⅡJ** 迅速结合到复合物上。这些蛋白是体外转录所必需的，并与复合体按一定次序结合。TFⅡH 是一个至少由五个亚基组成的大复合体。TFⅡJ 的特性还有待全面鉴定。

TFⅡH 作用下的 CTD 磷酸化

TFⅡH 是一个大的既有**激酶**活性又有**螺旋酶**活性的多组分蛋白复合体。TFⅡH 的活性在于导致 RNA 聚合酶的羧基末端结构域（CTD）的**磷酸化**（参见 M1）。这种磷酸化的结果导致了 RNA 聚合酶复合体的形成，并且允许 RNA 聚合酶离开启动子区域。因此 TFⅡH 看来在转录延伸的调控上具有非常重要的功能（参见 N2 中的 Tat 蛋白功能）。TFⅡH 的组成成分对于 DNA 修复以及调控细胞周期的细胞周期依赖性激酶复合体的磷酸化也很重要。

起始转录复合体

许多不含 TATA 框的 RNA 聚合酶Ⅱ含有与其起始位点重叠的起始元件。在这些启动子上，似乎 TBP 是通过另一可与起始元件结合的 DNA 结合蛋白而被添加到启动子上的，接着以一种类似于在含有 TATA 框的启动子上发生的方式来吸收别的转录因子和 RNA 聚合酶。

（郑　润译　李文君校）

N1 真核生物的转录因子

要 点

转录因子结构域结构

转录因子的分子结构由二个结构域组成，一部分用来结合 DNA，另一部分用来激活转录。一些转录因子还有二聚体结构域。

DNA 结合结构域

- 螺旋-转角-螺旋这样的结构域存在于原核生物的 DNA 结合蛋白如乳糖阻抑蛋白同源异型框序列编码的 60 个氨基酸的结构域中。两个 α-螺旋由一个特定角度的 β-转角分开，其中用于认别的 α-螺旋与 DNA 发生相互作用。
- 锌指结构域有二种，一种是通过二个半胱氨酸和两个组氨酸残基与锌离子结合形成的 C_2H_2 型，另一种是通过 4 个半胱氨酸残基与锌离子相连的 C_4 型。C_2H_2 型的锌指与 DNA 结合需要三个或更多的“手指”，而 C_4 型则成对出现，以二聚体的形式与 DNA 结合。
- 碱性结构域（basic domain）是与亮氨酸拉链以及螺旋-环-螺旋（helix-loop-helix，HLH）二聚体结构域相结合的，碱性结构域一般以二聚体的形式与 DNA 结合。

二聚体结构域

- 在亮氨酸拉链的 α-螺旋区内疏水亮氨酸精确地相距 7 个氨基酸残基，这样在 α-螺旋的某一侧面，每两圈就会出一个亮氨酸。两个单体的蛋白通过亮氨酸的相互作用实现二聚化。
- HLH 蛋白有两个 α-螺旋，中间被非螺旋的肽环隔开。C 端 α-螺旋某一面的疏水氨基酸残基可以进行二聚。和亮氨酸拉链一样，HLH 基序通常与为 DNA 二聚化所需的 N 端碱性结构域相邻。

转录激活结构域

- 许多转录激活因子都含有由高比例酸性氨基酸组成的酸性激活结构域。
- 富含谷氨酰胺结构域包含高比例的谷氨酰胺残基，存在于转录激活结构域中，例如转录因子 SP1。
- 富含脯氨酸结构域包含一段连续的脯氨酸残基，能够激活转录，原癌基因 *c-jun* 的产物中就存在一个富含脯氨酸的激活结构域。

阻抑物结构域	阻抑物可通过掩盖 DNA 结合区域或降低转录活性来间接地阻断转录因子的活动，或者他们本身就含有一个特定的直接阻抑的结构域。
转录调控的对象	起始转录复合物中不同的激活结构域可能有多个不同的对象，例如 TFⅡD 中的 $TAF_{Ⅱ}$、TFⅡB 和进行 C 端结构域的磷酸化的 THⅡH。
相关主题	蛋白质结构与功能（B2） 转录调控举例（N2） RNA 聚合酶Ⅱ基因：启动子与增强子（M4） 噬菌体（R2） 通用转录因子与 RNA 聚合酶 Ⅱ的起始（M5）

转录因子结构域结构

与起始复合物中的通用转录因子（general transcription factor）（参见 M5）不同，转录因子是通过它们与启动子的特定结构（如：上游调控元件或增强子区域）的亲和力才首次被发现的。这些因子有两种独特的活性：首先它们特异地与 DNA 结合位点相结合，然后激活转录。这些活性可以独立分配给特定的蛋白结构域，分别称作**激活结构域**和 **DNA 结合结构域**。另外，许多转录因子以同型或异型的**二聚体**形式存在。还有一些转录因子有**配体结合结构域**，可以与附加的小分子结合来调控转录因子自身的活性，类固醇激素受体（参见 N2）是包含所有这四种结构域的一个例子。

对酵母转录因子 **Ga14** 和 **Gcn4** 的诱变（mutagenesis）实验发现它们的 DNA 结合和转录激活结构域处在蛋白质的不同部位。通过实验把这些激活结构域融合到细菌的 **LexA 阻抑物**中，该杂交融合蛋白可从包含 *lexA* 操纵基因的启动子激活转录，表明酵母蛋白的转录激活功能是与其 DNA 结合活性相分离的。这类实验叫做**结构域交换实验**。

DNA 结合结构域

螺旋-转角-螺旋结构域

这类结构域的特点是它的 DNA 结合蛋白包含一个 60 个氨基酸的**同源域**（homeodomain），该同源域是由一段称为**同源框**（homeobox）序列编码的（参见 N2）。在果蝇的触角足（Antennapedia）转录因子中，该结构域是由 4 个 α-螺旋组成，螺旋Ⅱ与Ⅲ之间保持了适当的角度，并被一个颇具特征的 β-转角所分隔。这个特殊的螺旋-转角-螺旋-结构（图 N1.1）也在其他蛋白中出现，如 λ 噬菌体的 Cro 阻抑物（参见 R2）、乳糖和色氨酸阻抑物（参见 L1 和 L2）以及 cAMP 受体蛋白（参见 L1）。这种结构域的一个被称为识别的螺旋部分处在大沟中，并与 DNA 发生作用。两个同源域因子 Bicoid 和 Antennapedia 的识别螺旋可以互换，也就交换了它们的 DNA 结合特性。实际上只要改变识别螺旋上的一个氨基酸就会改变其与

DNA 的结合特性。

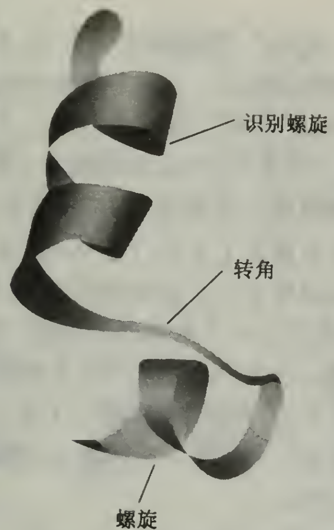


图 N1.1 螺旋-转角-螺旋的核心结构

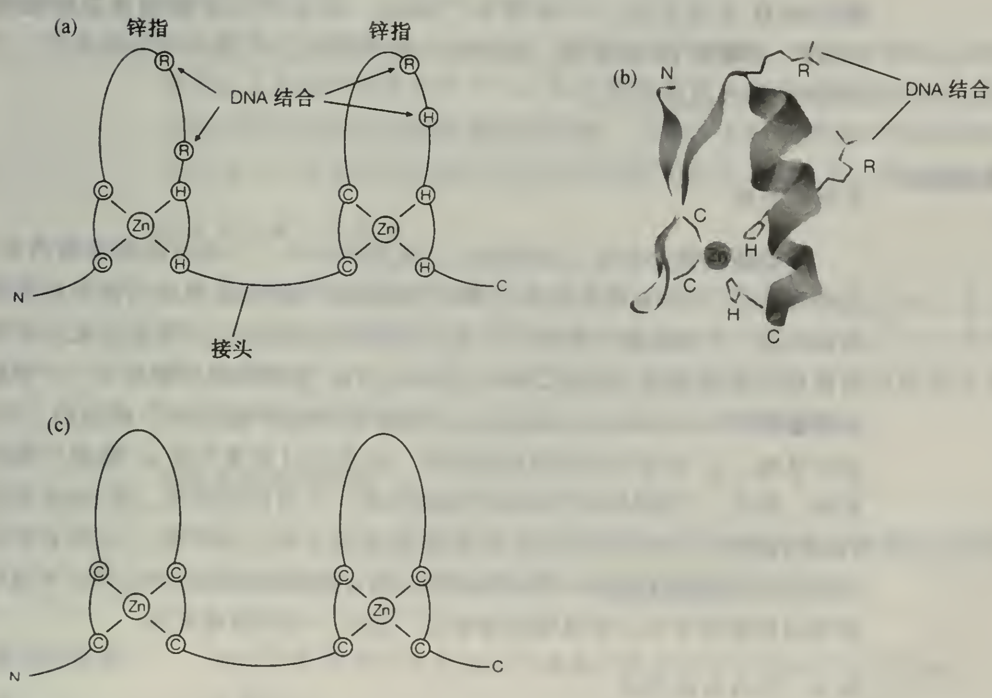


图 N1.2 (a) C_2H_2 锌指结构；(b) 锌指折叠结构；(c) C_4 锌指结构

锌指结构域

这种结构域有两种形式，**“C_2H_2”型锌指**有一个 12 个氨基酸的环，通过 2 个半胱氨酸和两个组氨酸残基固定，这四个残基与锌离子在空间上形成一个四面体结构（图 N1.2a）。这种锌指折叠形成一个紧密的结构，由二条 β 链和一个 α-螺旋组成，后者与 DNA 大沟结合（图 N1.2b）。该 α-螺旋区域上含有保守的碱性氨基酸，负责与 DNA 的结合。这个结构在 RNA 聚合酶Ⅲ转录因子 TFⅢA 中重复了 9 次（参见 M3），在转录因子 SP1 中也有三个拷贝。一般来说须要有三个或更多的 C_2H_2 型锌指才能与 DNA 结合。另一锌指结构是锌离子与 4 个半胱氨酸结合，它出现在一百多种类固醇激素受体转录因子中（参见 N2）。这些因子由同型或异型的二聚体组成，其中每一单体包含 2 个 C_4 锌指结构（图 N1.2c）。两个单体通过锌离子稳定折叠成更复杂的构象，再把每个单体的 α-螺旋插入到 DNA 的连续大沟中，这与 HLH 蛋白类似。

碱性结构域

在许多 DNA 结合蛋白中都发现了**碱性结构域**，它通常是与**亮氨酸拉链**或 **HLH** 基序中的一个联合在一起的，结果被称做**碱性亮氨酸拉链**（bZIP）或**碱性 HLH 蛋白**，蛋白的二聚作用使二个碱性结构域相邻，进而可与 DNA 发生作用。

二聚体结构域

亮氨酸拉链

亮氨酸拉链的肽链上每相隔七个残基就会有一个疏水的亮氨酸残基，这些残基位于 DNA 结合域的 C 端 α-螺旋上，这样 α-螺旋的侧面每两圈就会出现一个亮氨酸，形成一个疏水的表面。结果在 α-螺旋的疏水表面间就可以互相作用，形成二聚体（图 N1.3）。这种相互作用形成一个**卷曲的卷曲结构**（coiled-coil structure）。bZIP 转录因子包含一个碱性的 DNA 结合区域，其 N 端与亮氨酸拉链相连，这也可以被看作是 α-螺旋 C 端的延伸。每个 α-螺旋相连的碱性结构域形成一个对称的结构，沿 DNA 相反的方向延伸并与对称的 DNA 识别位点发生作用，最终像一个夹子夹在 DNA 上。亮氨酸拉链在一些利用 DNA 结合结构域的蛋白中也可以不通过碱性结构域而作为二聚体结构域使用，包括一些同源域蛋白。

螺旋-环-螺旋结构域

这一结构在总体上与亮氨酸拉链相似，只是它的二个 α-螺旋被一个非螺旋的多肽环分成二个单体蛋白，C 端 α-螺旋一侧的疏水残基可以二聚化，这种结构在 **MyoD** 这类蛋白中发现过（参见 N2）。与亮氨酸拉链一样，HLH 结构也经常与碱性结构域相邻，以形成 DNA 结合所需的二聚

体。这种异型二聚体的形成增加了转录因子所有组成成分的多样性和复杂性。

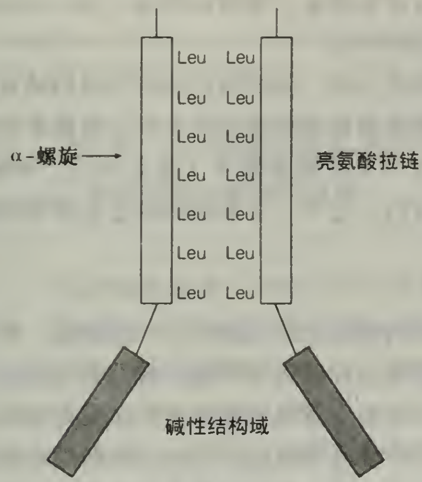


图 N1.3 bZIP 蛋白的亮氨酸拉链和碱性结构域二聚体

转录激活结构域

酸性激活结构域

通过比较酵母 Gcn4 和 Gal4 的转录激活结构域、哺乳动物糖皮质激素受体以及疱疹病毒激活子 VP16 发现它们都含有很高比例的酸性氨基酸，这样的结构域被称作**酸性激活结构域**，且是许多转录激活结构域的特征，但作为一个有效的转录激活结构域还需要其他什么特性仍不很清楚。

富含谷氨酰胺结构域

富含谷氨酰胺的结构域是在转录因子 SP1 的二个激活区域上首次发现的。与酸性结构域一样，谷氨酰胺残基所占的比例似乎比总体结构更重要。转录因子 SP1 中的富含谷氨酰胺转录激活结构域与果蝇的触足是蛋白间的结构域交换实验表明这些结构域是可以互相替换的。

富含脯氨酸结构域

在 c-Jun，AP2 和 Oct-2 等转录因子中所发现的**富含脯氨酸结构域**，与谷氨酰氨相似，有一个能激活转录的连续脯氨酸残基链。

阻抑物结构域

转录的阻抑有可能是通过间接地对激活因子功能的干扰而实现，有以下几种情况：

- 阻断了激活因子的 DNA 结合位点（与原核生物的阻抑蛋白一样，参见 L）。
- **非 DNA 结合复合体**的形成（如：类固醇激素受体的阻抑蛋白，或阻断

HLH 蛋白与 DNA 相互作用 Id 蛋白，因为它们缺少 DNA 结合的结构域，参见 N2)。

- 并非阻碍 DNA 结合而是掩盖了激活结构域（如：Gal80 掩盖了酵母转录因子 Gal4 的激活结构域)。

当然还存在一些其他方式，阻抑蛋白的特定结构域直接抑制转录。例如哺乳动物**甲状腺激素受体**的结构域在缺少甲状腺激素的情况下可以阻抑转录，而在与配体结合后又可激活转录（参见 N2)。**Wilms 癌基因** *WT1* 的产物就是一个抑癌蛋白，它有一个特定的缺乏电荷残基的富脯氨酸阻抑结构域。

转录调控的对象

激活结构域多样性的存在给我们提出了一个问题，即在起始转录复合体中它们的调控对象是相同的还是不同的？酸性激活结构域可以从下游的增强子位点激活转录，而富含脯氨酸结构域的激活力很弱，富含谷氨酰氨根本无法激活；在酵母中富含脯氨酸的结构域和酸性结构域具有活性，而富含谷氨酰胺的结构域则没有活性，这些都表明激活结构域有着不同的调控对象。不同的转录激活因子调控的对象可能有以下几种：

- 染色质结构；
- 通过特定的 TAF_{II} 与 TFⅡD 作用；
- 与 TFⅡB 作用；
- 对 TFⅡH 复合体的调整和作用导致 RNA 聚合酶 Ⅱ 的 CTD 的特异性磷酸化。

不同的激活结构域似乎有着不同的调控对象，而且转录起始和延伸过程的任何组分或阶段都可能成为调控的对象，从而实现转录的多阶段调控。

N2 转录调控举例

要 点

组成性转录因子：SP1	SP1 是一个广泛存在的转录因子，包含 3 个锌指结构和 2 个富含谷氨酰胺转录激活结构域。
激素调控：类固醇激素受体	类固醇激素进入细胞后与它的受体结合，致使该受体与被结合的抑制蛋白相分离，在二聚化后转移到核中。类固醇激素受体的 DNA 结合结构域与 DNA 上相应元件相结合，完成对目标基因的激活。甲状腺激素受体是一个转录阻抑蛋白，但它与甲状腺激素结合后，又可转为转录激活因子。
磷酸化调控：STAT 蛋白	γ 干扰素激活 JAK 激酶，后者可以使 STAT1α 发生磷酸化，STAT1α 二聚化后转移到细胞核中，进而激活目标基因表达。
转录延伸：HIV Tat	Tat 蛋白可以结合到 HIV RNA 的 5′端的一段称为 TAR 的序列上，如果没有 Tat 的结合，HIV 的转录就会过早终止。Tat-TAR 复合体在启动子处激活转录起始复合体中的 TFⅡH，导致 RNA 聚合酶Ⅱ羧基端结构域（CTD）的磷酸化，使聚合酶的全程转录成为可能。
细胞决定：*myoD*	*myoD* 及其相关基因（*myf5*、*mrf4* 和 *myogenin* ）的表达可以将非肌肉细胞转化成肌肉细胞。他们的表达激活特定的肌肉基因表达，并阻碍细胞的分裂。每一基因可以编码一种 HLH 转录因子，HLH 异型二聚体的形成使其结构具有多样性，并与非 DNA 结合抑制因子（non-DNA binding inhibitor）Id 的二聚化可以对这类转录因子进行调控。
胚胎发育：同源域蛋白	编码的一种 DNA 结合结构域的同源框最初是在果蝇中编码转录因子的同源异型基因中发现的。这些转录因子决定身体各部分的发育。果蝇和哺乳动物间的这些同源异型基因簇不管在功能还是在组成上都表现出很强的保守性，这说明他们在发育中扮演着很重要的角色。
相关主题	细胞分类（A1） 通用转录因子与 RNA 聚合酶Ⅱ的起始（M5） 真核生物的 DNA 复制（E4） 真核生物的转录因子（N1） 三种 RNA 聚合酶：性质与功能（M1）

组成性转录因子：SP1

SP1 与一段富含 GC 的保守序列 GGGCGG 相连，是一种组成性转录因子，其结合位点存在于许多持家基因的启动子中。SP1 存在于所有的细胞类型中，包含 3 个锌指结构以及 2 个富含谷氨酰胺转录激活结构域。SP1 的富含谷氨酰胺结构域与 $TAF_{II}110$ 发生特异性作用，$TAF_{II}110$ 是 TAF_{II} 中的一种，后者与 TATA 结合蛋白（TBP）相结合组成 TFⅡD。这就是 SP1 如何调控起始转录复合体的一种方式。

激素调控：类固醇激素受体

许多转录因子是由**激素**激活的。这些激素通常由一类细胞分泌，然后将信号转移给另一类细胞。**类固醇激素**是脂溶性的，可以穿过细胞膜与被称作**类固醇激素受体**的转录因子相互作用（参见 N1）。在没有类固醇激素存在的条件下，该受体与抑制蛋白结合，游离在细胞质中（图 N2.1）。当类固醇激素与受体结合后，可以使受体从抑制蛋白上游离出来，然后受体二聚化，进而转移到细胞核中。受体上的 DNA 结合结构域与特定的 DNA

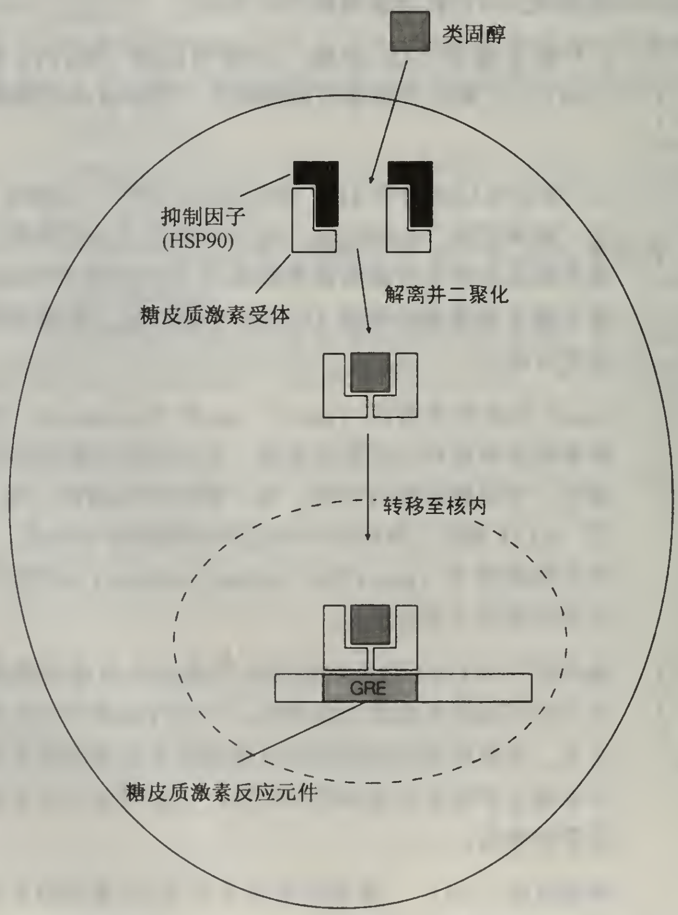
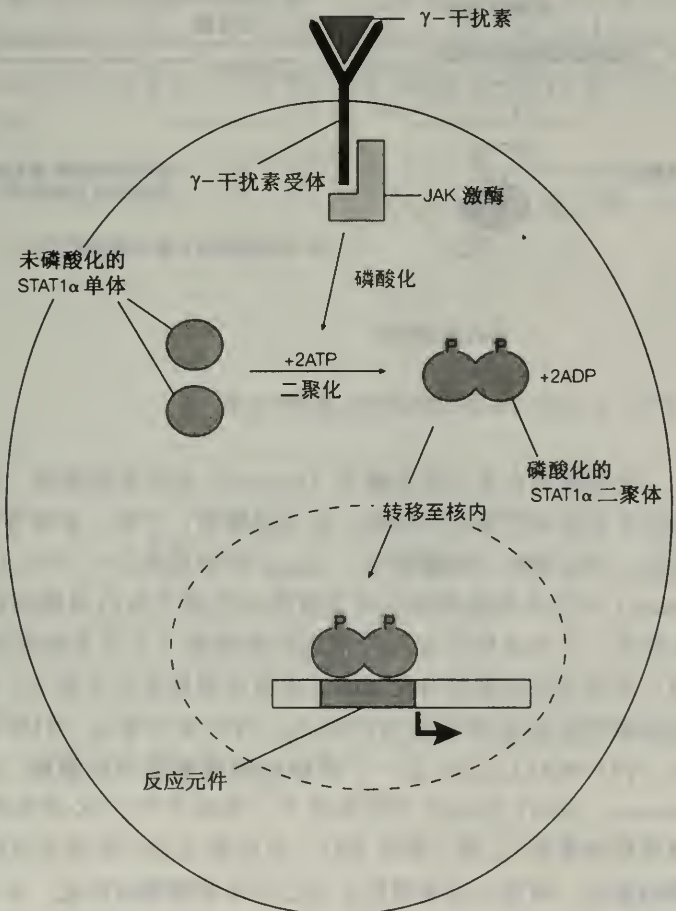


图 N2.1　类固醇激素激活糖皮质激素受体的模式图

序列或**反应元件**作用，从而激活目标基因。相关受体中比较重要的有：糖皮质激素、雌性激素、视黄酸和甲状腺激素受体。以上所述的一般模式并非对所有这些受体有效，例如甲状腺激素受体在没有激素结合的情况下是阻抑蛋白，但若与激素结合后又可以转化为转录激活因子。

磷酸化调控：STAT蛋白

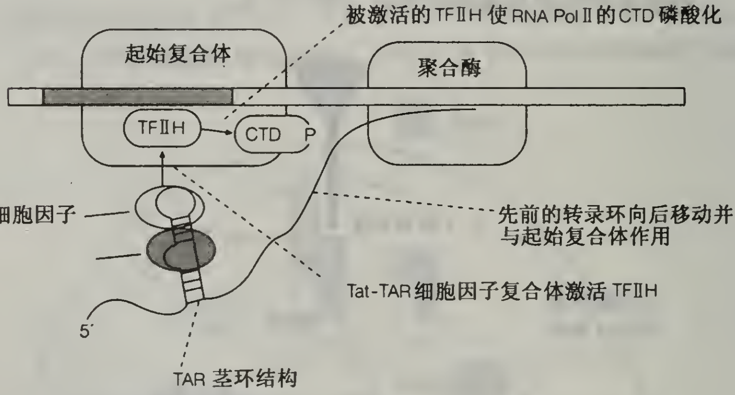
许多激素并不穿过细胞，它们与细胞表面的受体结合，通过称为**信号转导**的过程将信号传递给细胞内的蛋白，该过程通常涉及到蛋白磷酸化。γ-干扰素通过激活一种称为JAK的激酶诱发转录因子STAT1α的磷酸化。若STAT1α没有磷酸化，它只以单体的形成存在于细胞质中并且没有转录活性。然而它的一个特定酪氨酸残基发生磷酸化后，便能够形成同型二聚体，并从细胞质转移到细胞核中，进而激活在启动子处含有一保守DNA结合基序的目标基因的表达（图N2.2）。



图N2.2 由STAT1α转录因子的磷酸化及二聚化引起的γ-干扰素的转录激活过程

转录延伸：HIV Tat

人类免疫缺陷病毒（HIV）编码一种称为Tat的激活蛋白，该蛋白为HIV基因的大量表达所必需。Tat与RNA上的一段称为**TAR**的茎环结构结合。TAR是HIV RNA 5′端的转录起始点后的一段不翻译区域。在哺乳动物细胞中Tat所起的主要作用表现在转录延伸的过程中，若没有Tat的存在，RNA聚合酶Ⅱ转录复合体将因进程过慢而使HIV的转录过早终止。Tat可与RNA结合因子一起以复合体形式结合在转录物的TAR序列上，该蛋白-RNA复合体可以向后成环，并与装配在启动子处的新形成的转录起始复合体作用，这种作用导致TFⅡH的激酶活性被激活，结果RNA聚合酶Ⅱ的羧基端结构域（CTD）实现磷酸化，使得RNA聚合酶前进（参见M1和M5），使其完成HIV转录单位的阅读，最终实现HIV蛋白的大量合成（图N2.3）。



图N2.3 HIV Tat蛋白激活转录延伸的机制

细胞决定：*myoD*

肌肉细胞由来于称为**体节**（somites）的中胚层细胞。在机体能够通过细胞分化形成骨骼肌肉细胞（称为**生肌节**）之前，由体节负责肌肉细胞的生成，该过程称为**细胞决定**。*myoD*最初是作为一个在（**成肌细胞**，myoblast）未分化细胞中表达并负责肌肉形成与执行细胞决定的基因被鉴定出来的，它的过量表达可以使成纤维细胞（在许多组织中的基质铺垫细胞）变成表达肌细胞特异基因的类肌肉细胞和类生肌节。MyoD蛋白可以直接激活肌细胞特异基因的表达，同时也可激活*P21WAF1/CIP1*的表达。P21 WAF1/CIP1是一个**依赖细胞周期蛋白的激酶**（cyclin-dependent kinases，CDK）的小分子抑制因子，该因子对CDK活性的抑制可以使细胞周期停滞在G_1期（参见E3）。所以说*myoD*的表达是细胞退出细胞周期的原因，而退出细胞周期正是已分化肌细胞的特征。现在已发现的四种基因*myoD*、***myogenin***、***myf5***和***mrf4***中的任一基因的表达都可使成纤维细胞转化为肌细胞，它们所编码的蛋白都属于HLH转录因子家族的成员。MyoD与组成性HLH转录因子E12和E47结合成异型二聚体时活性最强。通过不同的组合，这些HLH型的转录因子能够产生多种各自具有不同功

能与作用的同型或异型的转录因子，这些因子可以受一种被称为**Id**的抑制因子调控。该 Id 缺乏与 DNA 结合的结构域，但都含有 HLH 的二聚化结构域，当它与 MyoD 和相关蛋白结合后，形成的异型二聚无法与 DNA 结合，因此也就无法调控转录。

胚胎发育：同源域蛋白

同源框是一段保守的 DNA 序列，编码一种称做同源异型域的螺旋-转角-螺旋的 DNA 结合蛋白（参见 N1）。同源异型域最初是在果蝇**同源异型基因**编码的转录因子中发现，该基因负责身体各部分的正确分化（参见 A1）。例如，同源异型基因中的一种 *Antennapedia* 的突变可使果蝇在应该长触角的地方长出腿来。这些基因对胚胎的空间构型的形成非常重要。同源框序列在真核生物的许多种中非常保守，含有同源框的基因在哺乳动物的发育中起着重要作用。在果蝇和哺乳动物中，同源异源基因被排列在同源基因相同顺序的基因簇中，这些基因同系物在胚胎的前后轴间也以相似的顺序进行表达。以上现象说明由保守同源异型基因所编码的 DNA 结合结构域是转录因子的特征，而转录因子在胚胎发育的过程中有着固定的功能。

（江　乐　译　刘进元　校）

O1 rRNA 加工与核糖体

要 点

RNA 的加工类型

无论在原核还是真核生物中，初生 RNA 转录物经过多种多样的加工，成为成熟的 RNA。三种最通常的方式是：①核酸酶剪切核苷酸；②向 5′端或 3′端添加核苷酸；③对核苷酸碱基或糖苷的修饰。

原核生物的 rRNA 加工

大肠杆菌最初的 30S 转录物是由 RNA 聚合酶转录 rRNA 七个操纵子之一而来。每个转录物包含一个拷贝的 5S rRNA、16S rRNA 和 23S rRNA 编码区以及一些 tRNA 序列。这个 6000nt 的转录物经过折叠、与蛋白复合、甲基化、再经过某些特殊酶（RNaseⅢ、M5、M16 和 M23）的酶切后，释放出成熟的 rRNA。

真核生物 rRNA 的加工

在真核生物的细胞核中，由 RNA 聚合酶Ⅰ（RNA Pol Ⅰ）转录 rRNA 基因，这些基因通常一前一后地重复排列，生成的较长单链前 rRNA 包含 18S、5.8S 和 28S 各一个拷贝。一系列特定的剪切使得不同的间隔区域从长链前 RNA 分子上分离出来。许多特异性核糖甲基化受小核糖核蛋白微粒（snRNP）控制，成熟的 rRNA 分子折叠并与核糖体蛋白复合。RNA 聚合酶Ⅲ从非相连基因合成 5S rRNA，5S rRNA 几乎不需要经过加工。

核糖核蛋白复合体及其研究

细胞中包含多种核糖核蛋白复合体（RNP），可以通过各种技术了解它们的结构和功能，技术包括解离、重新装配、电子显微镜观察、抗体的使用、RNA 酶的保护、RNA 结合、交联以及中子或 X 射线衍射。一些 RNP 的结构和功能已得到很好的鉴别。

原核生物的核糖体

核糖体是 rRNA 分子和特定的核糖体蛋白的复合体，这样大的 RNP 是细胞进行翻译的工厂。大肠杆菌 70S 核糖体由一个 50S 大亚基和一个 30S 小亚基组成，大亚基包含 31 种不同蛋白以及 5S rRNA、23S rRNA，小亚基包含 16S rRNA 分子和 21 种不同蛋白。

真核生物的核糖体

真核生物的核糖体比原核生物的更大更复杂，但是执行相同的功能。完整哺乳动物的 80S 核糖体由一个 60S 大亚基和一个 40S 小亚基组成。40S 小亚基包含一个 18S rRNA 分子和大约 30 种蛋白，而 60S 大亚基包含 5S rRNA、5.8S rRNA 和 28S rRNA 各一个以及大约 45 种蛋白。

相关主题	转录的基本原则（K1） RNA 聚合酶 I 基因：核糖体重复（M2） RNA 聚合酶Ⅲ基因：5S 基因与 tRNA 基因的转录（M3） tRNA 的加工、RNase P 和核酶（O2）

RNA 的加工类型

成熟的 RNA 分子很少是直接转录形成的（参见 K 和 M），大多数新生 RNA 分子经历了多种修饰才成为成熟的产物，**RNA 加工**是对**初生转录物**进行修饰的总称，最常见的加工类型包括：

1）内切核酸酶和外切酶对**核苷酸的切除**；

2）向初生 RNA 转录物或剪切产物的 3′端和 5′**端添加核苷酸**；

3）对某些特殊的核苷酸碱基或糖苷进行**修饰**。

这些加工发生在真核和原核生物中 RNA 的主要类型上。

原核生物的 rRNA 加工

在大肠杆菌中，有七种不同的 rRNA 操纵子分散在整个基因组中，分别被称为 *rrnH*、*rrnE* 等。每一操纵子均包含 **5S rRNA**、**16S rRNA**、**23S rRNA** 序列各一个，还包含一到四个 tRNA 的编码序列。这些初生转录物经过加工生成 rRNA 和 tRNA 分子，初生转录物的沉降系数为 30S（约 6000nt），但通常存在时间很短（图 O1.1a）。但在大肠杆菌 **RNA 酶Ⅲ**的缺陷型中，这些 30S 转录物有积累，表明 RNA 酶Ⅲ参与了 rRNA 的加工过程。其他酶缺陷型的研究表明大肠杆菌的 **M5**、**M16**、**M23** 等也参与了 rRNA 的加工过程。

大肠肝菌 rRNA 转录后加工经历了以下一系列特定步骤（图 O1.1a）：在 6000nt 的初生转录物转录之后甚至在转录过程中，转录物通过内部互补序列间的碱基配对折叠形成一些茎环结构。这些茎环状二级结构的形成使得一些蛋白与之结合形成**核糖核蛋白复合体**（RNP）。许多这样的蛋白会保持与 RNA 的结合状态并成为核糖体的一部分。在结合完成后，一些修饰就开始了，例如 24 种特定碱基的甲基化。*S*-腺苷甲硫氨酸（*S*-adenosylnethionrne，SAM）是一种常用的甲基化试剂，通常是将甲基化基团加在腺嘌呤碱基上，并随之发生最初的剪切，主要是由 RNA 酶Ⅲ进行剪切并释放出 5S、16S、23S 前体分子。随后 RNA 酶 M5、M16、M23 分别在每一前体分子的 5′端和 3′端进行进一步的剪切，在第二步剪切结束后释放出成熟的全长 RNA 分子。

真核生物 rRNA 的加工

真核生物中的 rRNA 同样也是由一个长的单链前体分子经过特定的修饰和剪切步骤后形成，但该过程还不是十分清楚。在许多真核生物中，rRNA 基因呈串联重复排列，包含 100 个或更多的转录单位，正如 M2 中所描述的那样，它们在核中由 RNA 聚合酶 I 转录（参见 D4）。在各种生物中前体有特定的大小，酵母菌中为 7000nt，哺乳动物中为 13500nt（图

O1.1b)。前体含 18S、5.8S 及 28S 编码区域的各一个拷贝，后两者合起来相当于原核生物中的 23S rRNA。真核生物中的 5S rRNA 由 RNA 聚合酶Ⅲ转录散在基因产生出 121nt 转录物，而且几乎不需要加工。

对于哺乳动物的前 rRNA，这个 13500nt（47S）的前体经历了一系列剪切（图 O1.1b），最初切除掉**外部转录间隔区**（ETS）1 和 2，再切去**内部转录间隔区**（ITS），然后释放出 32S 和 20S 的两个前 RNA，这两个片段都须经过进一步地修剪。在成熟的 RNA 分子产生前，5.8S 片段必须与 28S rRNA 完成碱基配对。真核生物的前 rRNA 的折叠和与蛋白复合都是与转录同时发生的，这些过程均在核中进行。甲基化大约在一百多个位点上进行，使之形成 2′-*O*-甲基核糖，现在已知**甲基化**是由聚集在一起的一系列核内**小分子核糖核蛋白微粒**在核内进行的。这些蛋白微粒在核中很丰富，他们的 snRNA 上包含一小段可与 rRNA 的一部分序列互补的序

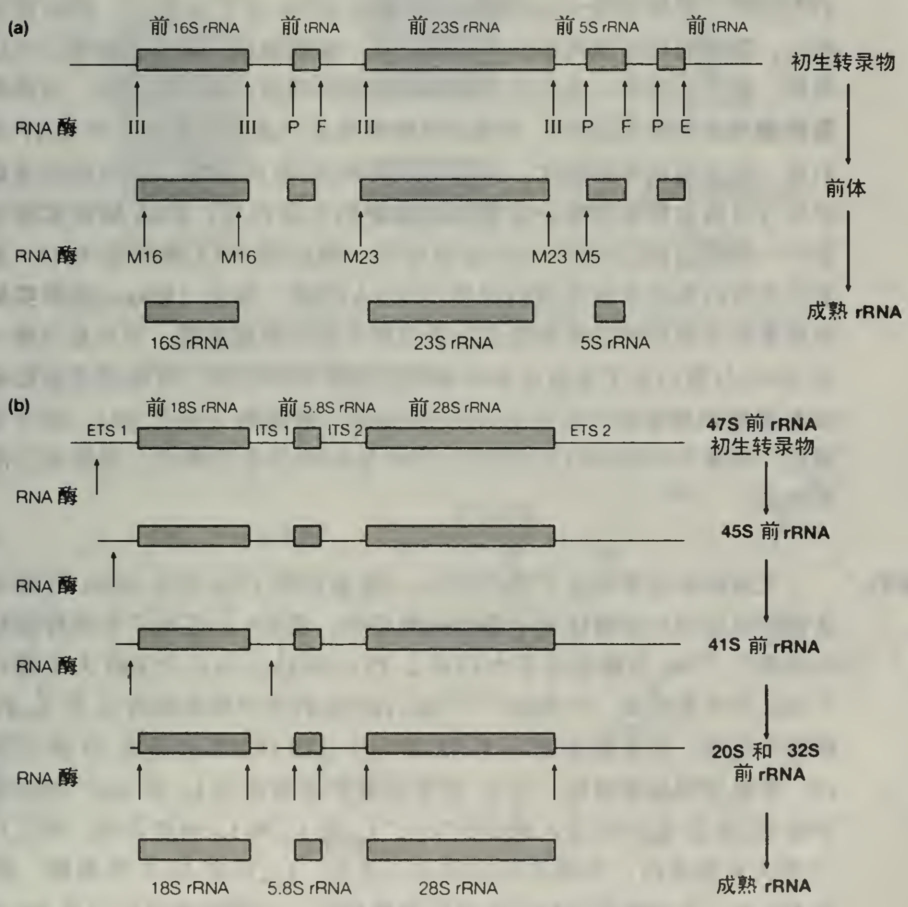


图 O1.1 (a) *E. coli* rRNA 初生转录物的加工；(b) 哺乳动物前rRNA 分子加工

列，通过配对决定甲基化修饰的部位。

至少有一种真核生物，嗜热四膜虫（*Tetrahymena thermophila*），其 rRNA 在具备功能前经历了一种不同寻常的加工过程。该前体包含一个在加工中必须被切除的内含子（参见 O3），虽然这种加工过程是在蛋白存在的情况下在胞内发生的，但在完全无蛋白存在的试管中，这个内含子居然还是被自身切除了。这种 RNA 可以折叠成有酶切活性的构型或者称作**核酶**（参见 O2）去完成自身剪切与连接。

核糖核蛋白复合体及其研究

在细胞内，RNA 分子通常与蛋白复合存在，特定的蛋白与特定的 RNA 结合。这种 RNA 与蛋白复合体称为**核糖核蛋白**（ribonucleoprotein, RNP)。**核糖体**是最大且最复杂的 RNP，在加工过程中由 rRNA 和特定的核糖体蛋白复合而成。其他的 RNP 将在 O2 和 O3 中讨论。用于研究 RNP 的方法有：分离可将 RNP 纯化后**分解**为 RNA 分子与蛋白，然后再分别鉴定；**重装配**用来发现组分相应结合在一起的规则，如果这些组分可以被修饰，就可以获得一些关于它们各自功能的线索；如果足够大，可用**电子显微镜**来直接观察 RNP，否则也可粗略地显示其总体形状；用 RNP 或他们单一组分的抗体来纯化，也可用来封闭 RNP 的功能，若与电子显微镜联用可以确定特定的组分在整个结构中的大致位置；**RNA 结合实验**可以显示一种特定蛋白与某一 RNA 的结合，随后用 RNA 酶处理 RNA-蛋白复合体可以显示出结合蛋白保护了 RNA 的哪一部分（**RNase 保护实验**）；运用紫外光照射经过和未经过化学试剂交联的**交联实验**，可以显示哪一部分 RNA 与蛋白分子在复合体中是紧密结合在一起的；而物理方法如**中子**和 **X 射线衍射**最终可以给出 RNP 完整的三维结构（参见 B3）。在本节和随后一些章节中所述的许多有关 RNP 结构的信息是通过上述研究方法所获得的。

原核生物的核糖体

大肠杆菌核糖体占干重的 25%（总蛋白的 10%和总 RNA 的 80%），这很好地体现出核糖体对于细胞的重要性。图 O1.2 显示了大肠杆菌核糖体的组分，70S 核糖体分子大约有 2.75×10^6Da，由一个 **50S** 大亚基和一个 **30S** 小亚基组成。后者由一个 16S rRNA 的分子和被标为 S_1 到 S_{21}的 21 种蛋白组成。大亚基包含一个 5S 和一个 23S rRNA 分子及 31 种不同蛋白，根据双向凝胶电泳，这 31 种蛋白被分别命名为 L_1 到 L_{34}，然而随后发现 L_{26}就是 S_{20}，L_7是乙酰化的 L_{12}，L_8 是 L_{10}和 L_7 的复合体，所以只有 31 种大亚基蛋白。核糖体蛋白的大小各异，L_{34}只有 46 个氨基酸，而 S_1 有 557 个。大多数相对较小的蛋白是碱性的，这可能与它们结合 RNA 有关。用 RNA 和蛋白组分可重新装配成有功能的大肠杆菌核糖体，并且已形成一条固定的装配路线。如图 O1.3 所示，多种方法研究已揭示出大肠杆菌核糖亚基的结构。

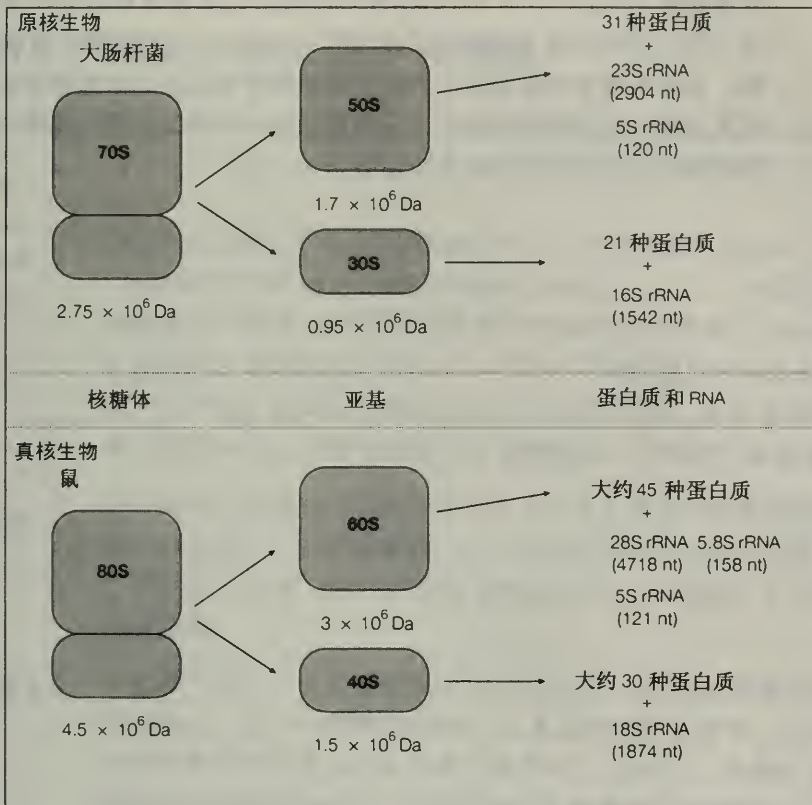


图 O1.2 典型的原核与真核糖体的组成



图 O1.3 *E. coli* 核糖体特征。(a) 30S 亚基；(b) 50S 亚基；(c) 完整的 70S 核糖体

真核生物的核糖体

图 O1.2 显示了真核生物大鼠核糖体 80S 的相应大小及组成成分。在 60S 大亚基中，包含了大约 45 种蛋白、一个 5S rRNA 分子、一个 5.8S rRNA 及一个 28S rRNA 分子，后两者相当于原核生物 23S rRNA。40S 小

亚基包含 18S rRNA 和大约 30 种不同蛋白。虽然真核生物中的 rRNA 均大一些，但在相当大程度上均保留了各自 RNA 分子的二级结构的保守性，由于其更大的复杂性，真核生物核糖体亚基还无法重新装备成有功能的复合体，它们的结构也不十分清楚。在一个典型的真核细胞中的核糖体每秒总共可以生成约 100 万个肽键。

O2 tRNA 的加工、RNA 酶 P 和核酶

要 点

原核生物的 tRNA 加工	经对较长的前 tRNA 转录物进行加工后形成成熟的 tRNA。该加工过程涉及特定的内切和外切酶如 RNase D、E、F、P 剪切以及随后对每一特定 tRNA 类型所进行的独特碱基修饰。RNase E、F 开始对 3′端切割后，RNase D 可以对 3′端修剪直至其比成熟长度多 2 个 nt。RNase P 酶然后对 5′端进行切除，生成成熟的 5′端。最后 RNase D 将移去两个 3′端的残基，同时进行碱基修饰。
真核生物的 tRNA 加工	许多真核生物前 tRNA 会含有内含子以及 5′端和 3′端额外的核苷酸，这些在加工中必须去除。与原核生物 tRNA 不同，真核生物 3′端的 CCA 是由 tRNA 核酸转移酶加上的，并且发生了许多的碱基修饰。
核糖核酸酶 P	在大肠杆菌中，由一个 377nt RNA 和一个 13.7kDa 的蛋白组成的 RNase P 是一个简单的 RNP。在真核和原核生物中，该 RNP 的功能都是剪切前 tRNA 5′端先导序列。在体外，单独的 RNA 成分就可剪切前 tRNA，所以它是一个有催化功能的 RNA 或称核酶。
核 酶	数个生化反应可由 RNA 酶或核酶催化，这种具催化功能的 RNA 可以剪切自身或其他的 RNA 分子，或者完成连接或自身剪接反应。核酶可以独自进行催化，但在体内通常与蛋白结合在一起工作，这将提高它们的催化活力。科学家们现在可以设计核酶作为剪切 RNA 工具。
相关主题	RNA 聚合酶Ⅲ基因：5S 基因与 tRNA 基因的转录（M3） tRNA 的结构与功能（P2） rRNA 加工与核糖体（O1）

原核生物的 tRNA 加工

从 O1 的图 O1.1 中可知大肠杆菌 rRNA 操纵子中包含了 tRNA 编码区域，除此之外，大肠杆菌还有其他包含多至七种 tRNA 基因的操纵子，这些 tRNA 基因被间隔序列相隔。成熟的 tRNA 分子可以从这些操纵子的前体转录物加工，即经过 **RNase E、F、D、P** 按顺序的一系列加工步骤，图 O2.1 显示了大肠杆菌 $tRNA^{Tyr}$的加工过程。一旦转录物前体经过折叠形成特征茎环结构（参见 P2），内切核酸酶（参见 D2）（RNase E、F）在 3′端切除一个侧翼序列，在前体的 3′端留下一个额外的九核苷酸，然后外

切酶 RNase D 依次切去 3′端区域的 7 个核苷酸，接着 RNase P 进行内切产生出成熟的 tRNA 5′端。随后，RNase D 再继续除去 3′端剩余的两个核苷酸，产生出成熟的 3′端。最后 tRNA 再经过一系列碱基修饰。不同的前 tRNA 以相似的方式进行加工，但**碱基修饰**对于各种特定 RNA 类型都是独特的。最通常的 tRNA 修饰参见 P2 的图 P2.1。

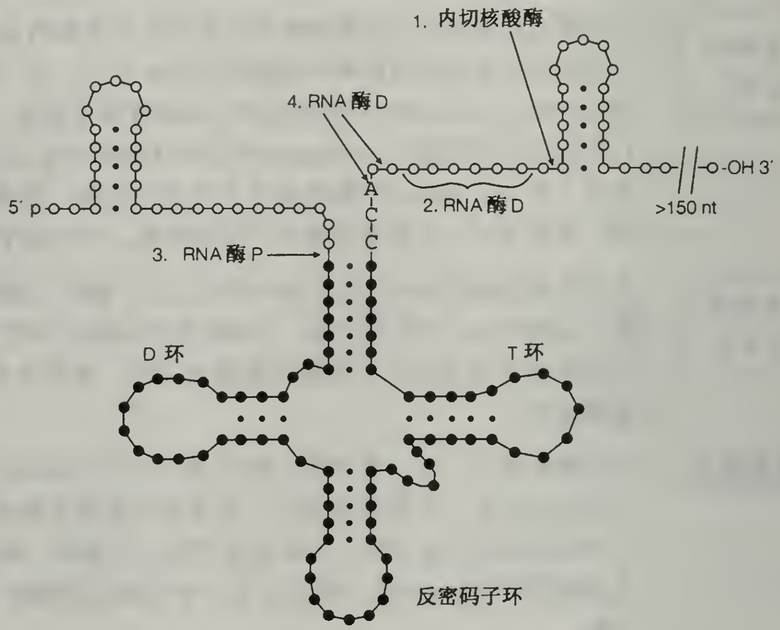


图 O2.1 *E. coli* 中前 tRNA 的加工过程

真核生物的 tRNA 加工

真核生物酵母菌 tRNA^{Tyr}的加工过程以图 O2.2 表示。如图所示该**前 tRNA** 带有一个 16nt 5′端前导区，一个 14nt **内含子**和 2 个额外的 3′端核苷酸。初生的转录物形成一个具特征茎环的二级结构，内切酶正是识别这种结构并切除 5′端前导区和 2 个额外的 3′端核苷酸。真核生物与原核生物的主要不同点是原核生物成熟 tRNA 3′端的 CCA 是基因编码的，但真核生物核编码 tRNA 不属这种情况。当 3′端的两个核苷酸被切除后，**tRNA 核苷酸转移酶**将 5′- CCA - 3′序列添加到 tRNA 的 3′端，产生出成熟的 tRNA 3′端。下一步是切除内含子，由内切酶将内含子两端切除后，再将两个半个的 tRNA 分子连接在一起。酵母菌前 tRNA 中的内含子可在脊椎动物细胞内加工，所以说真核生物 tRNA 加工机制看来在进化中是高度保守的。

核糖核酸酶 P

RNase P 是由一个 RNA 分子和一个蛋白分子组成的一种内切酶，是一种很简单的 RNP。该酶在细胞中的作用是修剪前 tRNA 分子产生出成熟的 5′端。RNase P 在真核和原核生物中都有发现。真核生物中 RNase P

位于核内，是一种核内小分子 RNP（snRNP）。在大肠杆菌中该内切核酸酶由一个 377nt RNA 和一个 13.7kDa 小的碱性蛋白组成，该 RNA 的二级结构在进化中高度保守。令人惊奇的是如果向试管中加入前 tRNA，单独的 RNA 组分就可行使内切核酸酶功能。这种 RNA 是一种具催化活性的 RNA，或称为**核酶**，即使无蛋白存在也可催化化学反应。现在已知有多种核酶，体外的 RNase P 催化反应比在体内需更高的镁离子浓度，因此在细胞内蛋白组分可能有助于催化反应。

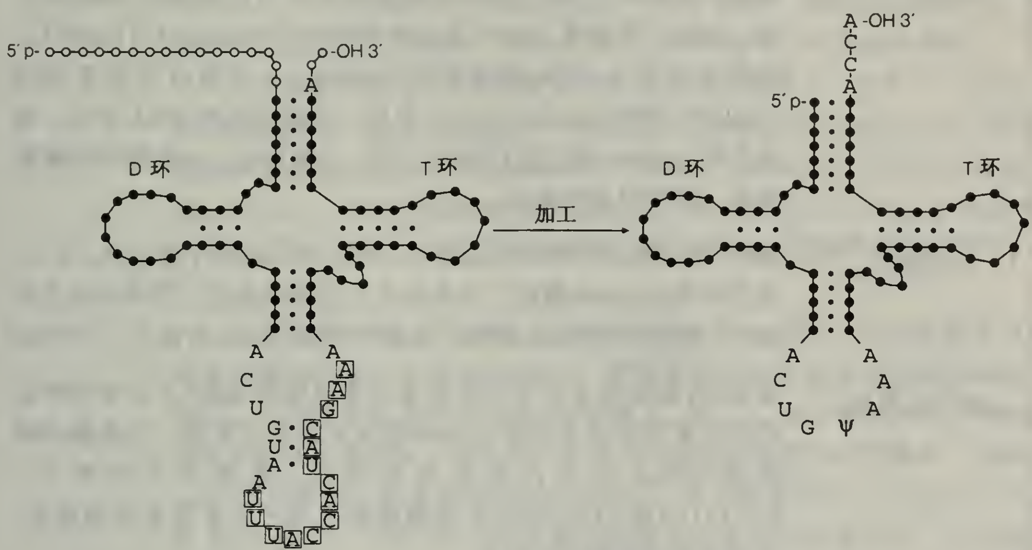


图 O2.2 酵母菌前 tRNA^{Tyr}的加工过程，用方框表示内含子核苷酸

核酶 核酶是一种可以催化特定生化反应的 RNA 分子。虽然发现较晚，但已知在自然界有几种核酶。目前研究者已能够用胞外选择技术来设计出新型的核酶。RNase P 是一种广泛存在可使 tRNA 成熟的核酶，它的 RNA 组分是一种可以做为内切核酸酶的核酶。在四膜虫（*Tetrahymena*）的 rRNA 大亚基中有一个内含子，可在无蛋白条件下，可以对转录物进行体外自我剪接（参见 O1）。这种加工称为**自我剪接**，该过程需要鸟苷或磷酸化的衍生物作为辅因子。体外反应的效率只有体内反应的五十分之一，所以细胞内蛋白可能会有助于体内的催化反应。在一些植物病毒复制过程中会产生出连体分子，这是由于在完成一个循环 RNA 合成后聚合酶继续合成所致。这种分子可以折叠然后自身剪切成为一个基因组大小长度的单体。对这些自身剪切分子的研究已经确定出所需最小序列，研究者们已能设计并合成出能以顺式或反式切割其他目标 RNA 分子的核酶。近来更多的研究转向通过利用核酶在体内剪切 mRNA 分子来抑制基因表达，也许这可能会阻止病毒复制，杀死癌细胞以及通过使基因失活来研究新基因的功能。

O3 mRNA 加工、hnRNP 和 snRNP

要 点

mRNA 的加工	原核生物的 mRNA 基本不需要加工，在转录结束前就可以开始翻译。在真核生物中，RNA 聚合酶Ⅱ负责较长的前 mRNA 合成，这些不同的前 mRNA 被统称为核内不均一 RNA（hnRNA）。特定蛋白与 hnRNA 结合形成 hnRNP，一些核内小分子 RNP（snRNP）颗粒与 hnRNP 发生作用，完成对某些 RNA 加工。真核生物 hnRNA 加工包括四个方面：5′端加帽、3′端剪切和聚腺苷化、剪接以及甲基化。
hnRNP	RNA 聚合酶Ⅱ转录物（hnRNA）与三种 hnRNA 蛋白 A、B、C 复合形成 hnRNP 颗粒，其中包含三个四聚体各三个拷贝和大约 600～700 核苷酸的 hnRNA。hnRNP 辅助 RNA 的加工。
snRNP 颗粒	由 RNA 聚合酶Ⅱ转录的许多富含尿嘧啶的 snRNA 分子与特定的蛋白复合形成 snRNP，其中绝大多数参与剪接，大多参与确定前 rRNA 分子内的甲基化位点，那些含有序列 5′-RA（U）$_n$GR-3′的 snRNP 在细胞质内与八个常见蛋白相结合，经甲基化后再被运回核中。
5′端加帽	当 RNA 聚合酶Ⅱ聚合的转录产物达到约 25nt 长时，在其 5′端加上了一个 7-甲基化鸟苷（m^7G），该 7-甲基化鸟苷或称帽子结构是以 5′-5′方向相连的，用来防止 5′-外切酶的攻击，但却有利于剪接、转运和翻译的进行。
3′端剪切及加尾	许多真核生物的前 mRNA 在多聚 A 位点被剪切后，多聚 A 聚合酶（RAP）在形成的 RNA 3′端加上约 250 nt 的 poly（A）尾巴，形成成熟的 3′端。
剪 接	在真核生物前 mRNA 加工过程中，打断编码区的内含子序列被切除，然后两侧的外显子片段相连接。这种剪接反应在核内进行，需要内含子有 5′-GU、AG-3′以及一段分支点序列。在两步反应中，内含子先以一个具尾的环状分子或套索状分子形式被删除，然后被降解。剪接包括 snRNP 与保守序列结合，形成剪接体，在其内发生剪切与连接反应。
前 mRNA 的甲基化	在前 mRNA 上的一小部分腺嘌呤残基，当序列为 5′-RRACX-3′时（其中 R 为嘌呤）会在 N6 位置发生甲基化。

相关主题	RNA 聚合酶Ⅱ基因：启动子与增强子（M4） rRNA 加工与核糖体（O1）

mRNA 的加工

原核生物中 mRNA 转录物根本不需要或很少需要加工。事实上核糖体在 mRNA 分子的合成未完成前就开始翻译。原核生物 mRNA 很快从 5′端的开始降解，所以第一个**顺反子**（蛋白编码区）的翻译被限制在很短的时间内。一些内部顺反子为保护自身，在 3′端和 5′端形成茎环结构，做为对外切酶暂时性的屏障，所以在其完全降解前可以增加翻译频率。

因为真核生物 RNA 聚合酶Ⅱ要转录基因种类很多、大小不一从 60～300nt 的核内小 RNA（snRNA）到长度超过 100kb 的 *Antennapedia* 基因。该酶转录产物的群体统称为**核内不均一 RNA**（hnRNA），那些即将被加工成 mRNA 分子的转录物被称为**前 mRNA**，前 mRNA 分子经过 5′端加帽、3′端剪切及加多聚 A 尾、剪接和甲基化加工产生出成熟的 mRNA 分子。

hnRNP

由 RNA 聚合酶Ⅱ合成的 hnRNA 主要是前 mRNA，并很快被蛋白包裹形成**核内不均一 RNP**（hnRNP），所涉及蛋白被分成 A～U 类的 hnRNP 蛋白。三种非常丰富的 hnRNP 蛋白是 A、B、C，每一种均有两种构型。从核中提纯这些物质，可得到一种约 30～40S 的同源性相当高的、被称为 hnRNP 的颗粒。这些颗粒直径约 20nm，每个颗粒包含一个 600～700nt 的 RNA 分子，并与三种不同四聚体的各三个拷贝形成复合体。这些四聚体是 $(A_1)_3B_2$、$(A_2)_3B_1$ 和 $(C_1)_3C_2$，hnRNP 蛋白被认为有助于保持 hnRNA的单链状态，并辅助各种 RNA 加工反应。

snRNP 颗粒

RNA 聚合酶Ⅱ也转录大部分的 **snRNA**，后者常与特定蛋白形成 snRNP，这些 RNA 富含尿嘧啶，因此被命名为 U1、U2 等。含量最丰富的是那些参与前 mRNA 剪接的 U1、U2、、U4、U5 和 U6。而且 snRNP 种类正在不断增加，大部分好像参与了前 rRNA 加工中甲基化位点的确定，因此位于核中（参见 O1）。主要核质 snRNP 由单一的 snRNA 和一组 8 个碱性蛋白及多种 snRNP 特定蛋白构成。这些核心蛋白，依据识别它们的抗体被命名为 **Sm 蛋白**，需要 RNA 单链区中有一段 $5'-RA(U)_nGR-3'$序列。U6 中不含该序列，但通常进行与含有该序列的 U4 碱基配对。snRNP 的构成过程如下：由 RNA 聚合酶Ⅱ在核中合成，并具有一个正常的 5′帽子结构（参见下文），然后进入细胞质，与核心蛋白以及其他特定蛋白结合，然后再转运核中行使剪接功能。在它们的 5′端帽结构上得到两个甲基。

5′端加帽

RNA 聚合酶开始转录后不久，在 RNA 链长度超过 20～30nt 之前，其 5′端经化学修饰被加上了一个 **7-甲基鸟苷**残基（图 O3.1），这种 5′端

的修饰被称为帽子结构。与正常的 3′-5′连接方式不同，它通过一个 GMP 核苷酸以反方向加到新生的 RNA 转录物上产生，从而形成 5′-5′三磷酸桥。催化这一添加核苷酸反应的酶是 mRNA 鸟苷转移酶，特别的是在脊椎动物中，它首先对糖苷甲基化，然后再转录核苷酸。这种帽子结构对 5′外切核酸酶形成屏障，起到稳定转录物的作用，同时对 mRNA 及其前体的其他反应如剪接、转运和翻译也十分重要。

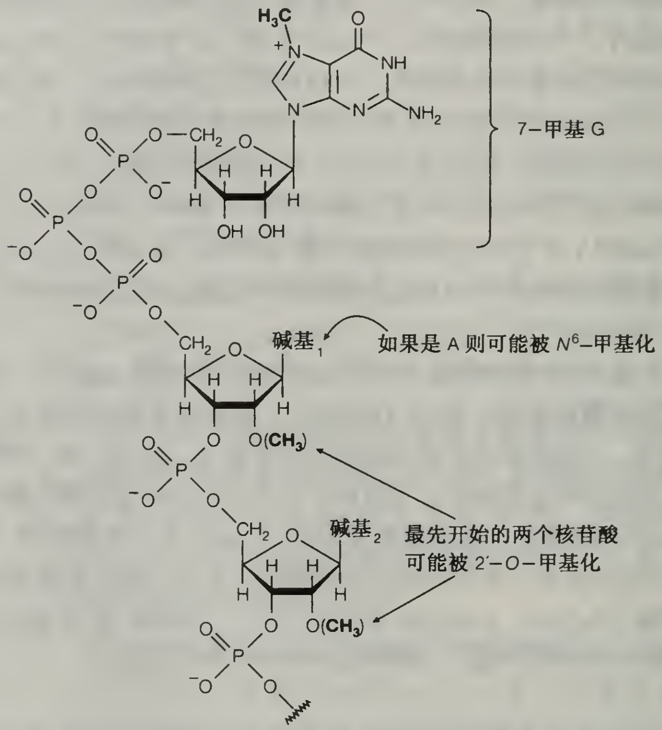


图 O3.1 真核生物 mRNA 的 5′端帽结构

3′端剪切及加尾

对于大多数前 mRNA 而言，其成熟的 mRNA 分子 3′端是经过剪切再加上一串多聚 A 残基即称为 **poly（A）尾**后而形成的。利用这一特征可将 mRNA 分子从其他类型 RNA 中纯化出来，用于 cDNA 文库的构建（参见 I1 和 I2），进而分离特定基因、分析它们的功能。

剪切和聚腺苷酸化反应需要 DNA 和前 mRNA 转录物上的特定序列，即 5′-AAUAAA-3′**聚腺苷酸化信号序列**，在其后 11～20nt 处紧随着一个“5′-YA-3′”结构，其中 Y 为嘧啶（图 O3.2a），在其下游常有一富含 GU 的序列。这些序列共同决定了**聚腺苷酸化位点**。

一些特定蛋白因子识别上述这些序列元件并与前 mRNA 结合，当复合体组装完成后即开始剪切。随后一种称为**聚腺苷酸聚合酶**（PAP）剪切后的前 mRNA 3′端加上多至 250 个 A 残基，即产生出 poly（A）尾部。前 mRNA上的 poly（A）尾被认为有助于稳定整个分子，因为有一种 **poly**

(A) **结合蛋白**与 poly (A) 结合，起抵抗 3′-外切核酸酶攻击的作用。另外，poly (A) 尾也有助于细胞质中成熟 mRNA 的翻译。组蛋白前 mRNA 不进行聚腺苷酸化，但也是在其一段特定序列被剪切后才生成成熟的 3′端。

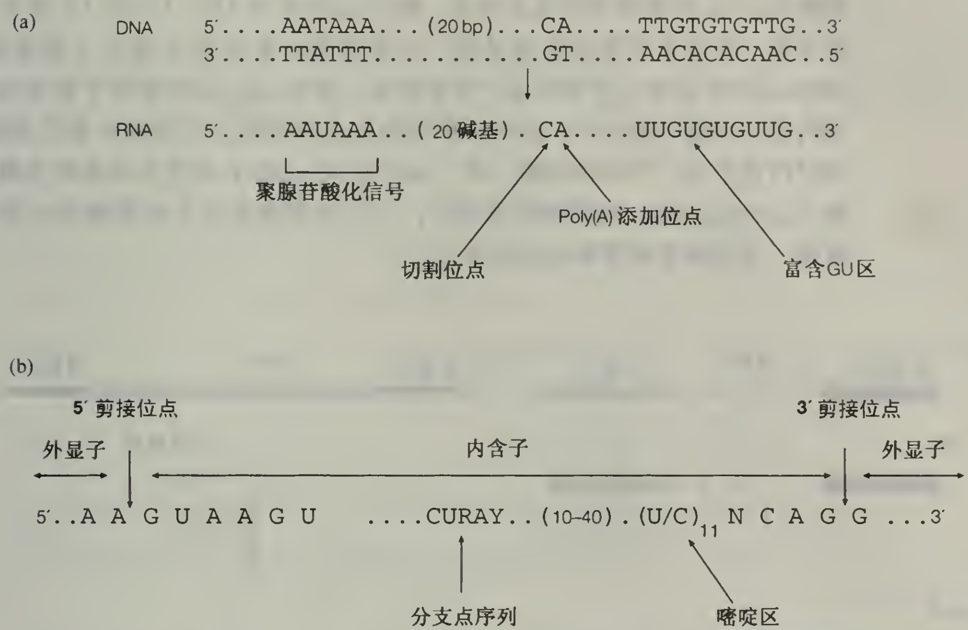


图 O3.2 典型的聚腺苷酸化位点序列 (a) 和剪接位点保守序列 (b)

剪 接

在真核生物前 mRNA 加工中，在聚腺苷酸化位点上游的一些已转录序列最终还是被切除了，以产生出成熟的 mRNA。这些序列从前 mRNA 的中心区切去后，两端相连成一体。这些插入序列或称**内含子**中断了成熟 mRNA 中相邻的**外显子**，即通常是 mRNA 的蛋白编码区。切除内含子将外显子连接在一起的过程称为**剪接**，就像 poly (A) 化的加工一样，发生在核内，且在成熟的 mRNA 分子转运到胞质之前。

剪接也需要有一组特殊序列（图 O3.2b）。几乎所有内含子的 5′端均有 5′-GU-3′序列，3′端通常是 5′-AG-3′序列，在 3′端的 AG 序列之前还有一段称做**嘧啶区**（ polypyrimidine tract）的富含嘧啶的序列。脊椎动物中在嘧啶区上游 10～40 个残基处有一被称为**分支点序列**的区域，5′-CURAY-3′序列（其中 R 代表嘌呤、Y 代表嘧啶），而在酵母中则是一个更特定的序列 5′-UACUAAC-3′。

剪接是一个**二步反应**（图 O3.3a），首先，在内含子 5′端 G 之前称为 **5′-剪接位点**处的被分支点序列中 A 残基上的 2′羟基攻击形成了一个称为**套索状**的尾环分子，并释放出外显子 1；接着在 **3′剪接位点**处的剪切发生 AG 中的 G 之后，随后两段外显子序列连接成一体。其内含子以套索形式

被释放，最终被降解。

剪接过程由 U1、U2、U4、U5 和 U6 **snRNP** 催化，也有其他剪接因子参与。这些 snRNP 中的 RNA 成分与 5′和 3′剪接点及分支点的多种保守碱基配对（图 O3.3b）。在剪接的前期，U1 snRNP 5′端与剪接位点的 5′端接合，U2 紧接着与分支点接合，随后 snRNP 中 U4、U5 和 U6 结合成复合体，使内含子形成环状结构，结果使得外显子的 3′端和 5′端接近。这些 snRNP 经相互作用形成一个复合体，使前 mRNA 折叠利于剪切的正确构型。该前 mRNA 与 snRNP 形成的复合体使得上下游的外显子靠近，同时内含子呈一个环状结构。这一 mRNA 与 snRNP 的复合体被称为**剪接体**（spliceosome）。在剪接体形成后，在二步剪接发生之前剪接体内进行重排，并以套索构型释放出内含子。

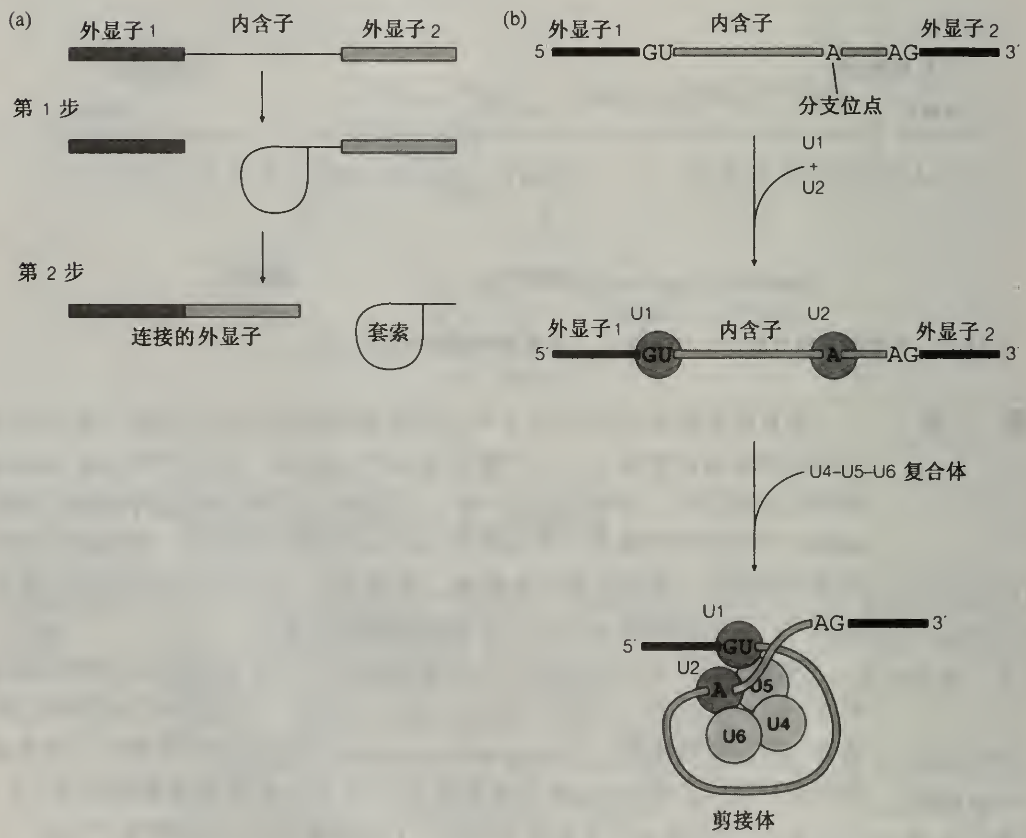


图 O3.3 真核生物前 mRNA 的剪接。(a) 两步反应；(b) snRNP 参与剪接体的形成

前 mRNA 的甲基化

许多前 mRNA 所经历的最终修饰或加工是某些碱基的甲基化。在脊椎动物中，最常见的是在 A 残基 N6 位置的甲基化修饰，特别是当 A 残基位于“5′- RRACX - 3′”序列中，这里的 X 很少是 G。多至 0.1% 的前

mRNA 中 A 残基是甲基化的，成熟的 mRNA 中甲基化修饰是高度保守，尽管甲基化的功能还不很清楚。

O4 可变 mRNA 加工

要点	
可变加工	可变 mRNA 加工可将前 mRNA 转化为一种以上的成熟 mRNA，通过使用不同的 poly（A）位点或不同的剪接方式实现对前 mRNA的加工。
可变 poly（A）位点	某些前 mRNA 含有一个以上的 poly（A）位点，可以在不同情况下（如在不同细胞类型中）产生出不同的成熟 mRNA。在某些情况下，相关因子将结合在某特定位点近旁，激活或抑制活性。
可变剪接	可以从一个特定基因转录物通过利用不同 5′、3′剪接位点产生出不同的成熟 mRNA。可变剪接有四种主要方式： 1）利用不同的启动子；　2）利用不同的 poly（A）位点； 3）保留某些内含子；　4）保留或去除某些外显子。 在不同细胞类型中会有不同的情况发生，这可能是由于细胞类型依赖因子在其激活或抑制结合位点附近的加工位点的活性所造成的。
RNA 编辑	RNA 加工的一种形式，是通过改变、插入或删除初生转录物特定部位的碱基而改变其中的核苷酸序列。例如人类 Apo-B 蛋白，在肠细胞内通过将 mRNA 上的一个 C 换为 U 产生出一个终止子，翻译形成一个截短蛋白。RNA 编辑会涉及向导 RNA（RNA 编辑的模板）。
相关主题	RNA 聚合酶 Ⅱ 基因：启动子与增强子（M4） mRNA 加工、hnRNP 和 snRNP（O3）

可变加工

在许多情况下真核生物的一个特定前 mRNA 会产生出一种以上的 mRNA。当特定外显子（**可变外显子**）被剪掉，在成熟的 mRNA 分子中不保留该外显子序列，就发生了可变加工。另外如有可被利用的可以选择 poly（A）位点，就会产生具不同 3′端的成熟 mRNA。**可变 RNA 加工**的类型包括**可变（或差异）剪接**和**可变（或差异）poly（A）加工。**

可变 poly（A）位点

一些前 mRNA 含有多于一组的用于剪切及加尾的序列（参见 O3）。细胞或生物体可以选择使用其中一组。若上游位点被使用，而位于被剪除部位的控制 mRNA 稳定性或定位的序列就有可能被切掉。因此可从同一基因产生出具有相同编码区，不同稳定性或定位的成熟 mRNA。在某些

情况下，在同一细胞中两个位点均可以一定频率（反映它们相对效率，强度）被利用，结果该细胞会同时含有两种 mRNA。poly（A）位点的效率反映出与保守序列的匹配程度（参见 O3）。另一方面，一种细胞可以只利用一个 poly（A）位点，而不同细胞采用另一 poly（A）位点。最可能的解释是在某一细胞中最强的位点被滥用，而在其他细胞中，由于存在某一可活化弱位点的因子，结果该位点被强制使用；或者存在一个阻止强位点被使用的因子。在某些情况下，可变 poly（A）位点的使用可导致采用不同的剪接方式（参见下文）。

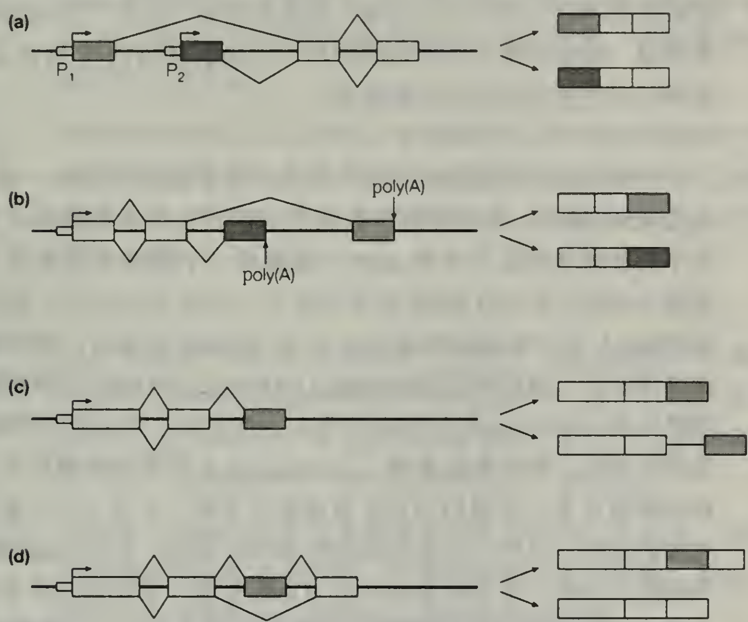


图 O4.1 可变剪接的方式。(a) 启动子 P_1 或 P_2 的选择；(b) 剪切和（或）多腺苷酸化位点的选择；(c) 内含子的保留；(d) 外显子的跳读。白盒表示外显子，灰盒表示可变外显子，细线表示内显子。根据 D. M. Freifelder (1987) 分子生物学，第二版绘制。

可变剪接

图 O4.1 概括出四种最常见的可变剪接方式。如图 O4.1a 所示启动子的选择决定剪接方式，这种情况在 α-淀粉酶和肌球蛋白的轻链基因中发生。从**上游启动子**转录的外显子有着较强的 5′剪接位点，结果“击败”从**下游启动子**的 5′位点，而获得第一个 3′剪接位点的使用。例如在唾液腺中，特定的转录因子使 α-淀粉酶基因从上游启动子开始转录；而在肝细胞中，使用下游启动子，结果较弱的（第二）5′剪接位点自然被使用。图 O4.1b 显示了 poly（A）位点的不同使用所导致的可变剪接。只有当下游的 poly（A）位点被使用，较强的 3′剪接位点才会存在，这样倒数第二个外显子将被删除；而当上游的 poly（A）位点被使用时（如不同细胞或发育的不同阶段），剪接便自然利用较弱的（上游）3′剪接位点。对于免

疫球蛋白来说，使用下游的 poly（A）位点可以使编码膜锚定区域的外显子保留下来，而使用上游位点这些区域就不存在，便产生分泌型的免疫球蛋白。在某些情况下，如图 O4.1c 所示内含子可被保留。若内含子包含一个终止密码子，翻译时就会产生一个截短蛋白，这可能导致产生一个没有活性的蛋白，正如果蝇体细胞的 P 元件转移酶的情况那样。在胚性细胞中，一个特定的因子可以使内含子正确剪接产生较长的 mRNA，进而翻译成功能酶，而体细胞中缺少这种特定因子。最后一种可变剪接（图 O4.1d）展示出某些外显子在不同的情况下可以被保留或被切除，其可能的原因是存在一种因子。在某一细胞类型中，该因子可以促进特定剪接位点使用，或者抑制另一位点的使用。大鼠肌钙蛋白 T 的前 mRNA 就是以这种不同的方式进行差异剪接的。

RNA 编辑

一种不同寻常的 RNA 加工模式，即原始转录物的一些序列被更改，叫做 **RNA 编辑**。有这样的一些例子，似乎在非脊椎动物中更加普遍。人体中的脱辅基脂蛋白 B 前 mRNA 编辑使一个碱基 C 变换为 U，结果在肠细胞 14500nt 分子的 6666 位点产生了一个终止密码子。在肝细胞中未编辑的 RNA 产生载脂蛋白 B100，一个 512kDa 的蛋白，而在肠细胞中，编辑结果产生了截短载脂蛋白 B48（241kDa）。相似地在神经细胞中，谷氨酸受体前 mRNA 上的一个 A 变为 G，产生出一种不同构型的受体。在纤毛原生动物，利什曼原虫属（*Leishmania*）中这种编辑更令人惊奇。当线粒体细胞色素 b 基因的 cDNA 被克隆出来时，它有一个与蛋白顺序相对应的编码区域，虽然已知它由线粒体基因组编码，然而在基因组中却没有发现相应的序列，最终获得的一些 cDNA 克隆，其序列是介于基因组序列与成熟 mRNA 之间的相应序列，好像原始转录物，在一些特定位点不断被编辑插入一些额外的 U，经过许多轮的编辑最终产生出可被翻译的成熟 RNA。一些被称为**向导 RNA** 的短 RNA 分子似乎参与其中，它们的序列与基因组 DNA 的序列及编辑后的 RNA 序列互补。一些其他类型的 RNA 编辑也有发现。

（江 乐 译 刘进元 校）

P1 遗传密码

要 点

特 性	遗传密码是核酸中的核苷酸序列指定蛋白质中的氨基酸序列的一种方式，是由三个核苷酸组成的三联体密码，密码子（三核苷酸组）是相连的（不重叠）、中间没有任何的停顿（没有逗号）。由于 64 个密码子中的许多是编码同一氨基酸的，因此遗传密码具简并性（有丰余）。
破 译	标准遗传密码是通过向具有有限翻译能力的细胞提取物中加入单一核苷酸的聚合物，多核苷酸共聚物或合成的核苷酸三联体而破译成功的，结果发现 61 种密码子编码 20 种氨基酸，而另外有三个终止密码子。
特 征	20 种氨基酸中的 18 种是由多个密码子（同义密码子）编码的，这些同义密码子在遗传密码表中成组排列，通常它们仅在密码子的第三位上有差异。第三位如果是嘧啶，那么它们总是编码同一氨基酸；如果是嘌呤，通常也是编码同一氨基酸。
突变的效应	成组同义密码子的存在意味着可将突变的效应最小化。密码子第三位的转换通常是无效的，多半的颠换也是无效的。在第一、二位的突变通常导致化学特性上相似类型氨基酸。
通 用 性	目前仍认为标准遗传密码是通用的，但在线粒体及一些单细胞生物中现已发现一些例外。
可 读 框	可读框（ORF）是指新测 DNA 序列中，由计算机辨认出的可能编码区域，它是从起始密码子起到终止密码子止的一段连续的密码子区域。
重叠基因	指一个基因的编码区部分或全部与另一基因的编码区重叠，因此一种可读框编码一种蛋白，而其他可能的可读框编码第二种蛋白的一部分或全部。某些较小病毒的基因组利用这种方式来增加它们基因组的编码能力。

相关主题

诱变（F1）　　真核生物蛋白质合成的起始（Q3）
tRNA 结构与功能（P2）　　翻译调控与翻译后加工（Q4）
蛋白质合成机制（Q2）

特性

遗传密码统一了核酸中四种核苷酸序列与蛋白质中 20 种氨基酸序列的对应性。研究证明遗传密码是三联体，每三个核苷酸编码一种氨基酸，这可以从数字上加以论征，由于$(4^2=16)<20<(4^3=64)$，可见指定每种氨基酸至少需要 3 个核苷酸。不过需要指定的氨基酸只有 20 种，而潜在的不同的三联体有 64 种，结果大多数氨基酸是由多于一个三联体所编码，因此说遗传密码是**简并的**，或有**丰余**。从固定的起始点开始，mRNA的编码区内每 3 个相邻的碱基代表一个密码子，被某一特定 tRNA 分子的一端的互补三联体（或称**反密码子**）所识别（参见 P2）。三联体密码是以线性方式被读，中间既不存在重叠也不存在停顿。一旦阅读自正确的起始点（参见 Q3）开始，密码子就以相邻三联体的形式被破译。随着越来越多的基因和蛋白序列信息的获得，越来越清楚地表明遗传密码几乎（而不是绝对）是普遍适用的，这有力地支持了所有生命都起源于共同的祖先的假说。

破译

20 世纪 60 年代，Nirenberg 建立了大肠杆菌**无细胞蛋白合成体系**。本质上这是经离心后的细胞裂解液，是用 DNase 处理以抑制任何转录，而在加入自然的或合成的 mRNA 后，又能进行有限的蛋白合成的体系。为了确认哪种氨基酸被聚合到新合成的多肽中，需平行进行 20 个反应。每个反应中含有 19 种非放射性的氨基酸和一种放射性标记的氨基酸。利用**多聚核苷酸磷酸化酶**合成只有一种核苷酸成分的 mRNA，即多聚尿嘧啶核苷酸［poly（U）］、多聚胞嘧啶核苷酸［poly（C）］，多聚腺嘌呤核苷酸［poly（A）］、以及多聚鸟嘌呤核苷酸［poly（G）］。如果加入一种这样**单一组成的合成 mRNA** 后，发生了蛋白质的合成，那么 20 个反应管中肯定会有一个试管，其放射活性将被渗入多肽中。用这种方法，结果发现 poly（U）指导了多聚苯丙氨酸的合成、poly（C）编码了多聚脯氨酸、poly（A）编码多聚赖氨酸，而 poly（G）没有蛋白的合成因为它形成了复杂的二级结构。

如果多聚核苷酸磷酸化酶能够被用来聚合两种核苷酸的混合物，比如说 U 和 G，以不同的比例如 0.76:0.24，那么在三联体中 GGG 出现最少，UUU 将是最普遍的，而三联体中含有 2 个 U 和一个 G 的出现频率则次之。利用这些**随机共聚物**作为合成 mRNA 加入到无细胞体系中，来确认某种特定氨基酸掺入到多肽中的比例，就可能确定出许多氨基酸的密码子的构成。但三联体密码子的精确顺序的确定还需要另外的信息。

直到 20 世纪 60 年代末，研究发现**合成的三核苷酸**能够附着在核糖体上，并与其对应的氨酰- tRNA 结合，而经膜过滤后，只有核糖体、合成的三联体以及氨酰- tRNA 形成的复合体（参见 P2）才能保留在膜上。若将含有所有氨酰- tRNA 的混合物分成 20 组进行实验，每组含有一种被放射性同位素标记的氨基酸，那么实验会显示出特定三联体可明确地来指定特定该氨基酸。结果总其有 61 种密码子编码氨基酸，而 3 个是终止密码

子（表 P1.1）（氨基酸的三字母代码和一字母代码参见 B1）。

表 P1.1 通用遗传密码

第一位	第二位								第三位
(5′端)	U		C		A		G		(3′端)
	Phe	UUU	Ser	UCU	Tyr	UAU	Cys	UGU	U
	Phe	UUC	Ser	UCC	Tyr	UAC	Cys	UGC	C
U	Leu	UUA	Ser	UCA	终止密码子	UAA	终止密码子	UGA	A
	Leu	UUG	Ser	UCG	终止密码子	UAG	Trp	UGG	G
	Leu	CUU	Pro	CCU	His	CAU	Arg	CGU	U
	Leu	CUC	Pro	CCC	His	CAC	Arg	CGC	C
C	Leu	CAU	Pro	CCA	Gln	CAA	Arg	CGA	A
	Leu	CUG	Pro	CCG	Gln	CAG	Arg	CGG	G
	Ile	AUU	Thr	ACU	Asn	AAU	Ser	AGU	U
	Ile	AUC	Thr	ACC	Asn	AAC	Ser	AGC	C
A	Ile	AUA	Thr	ACA	Lys	AAA	Arg	AGA	A
	Met	AUG	Thr	ACG	Lys	AAG	Arg	AGG	G
	Val	GUU	Ala	GCU	Asp	GAU	Gly	GGU	U
	Val	GUC	Ala	GCC	Asp	GAC	Gly	GGC	C
G	Val	GUA	Ala	GCA	Glu	GAA	Gly	GGA	A
	Val	GUG	Ala	GCG	Glu	GAG	Gly	GGG	G

特征

遗传密码具有简并性，即同一种氨基酸具有多个密码子，编码同一氨基酸的密码子被称为**同义密码子**。20 种氨基酸中，除了蛋氨酸和色氨酸仅有一个密码子外，其他 18 种均有一个以上的密码子。同义密码子不是随机排列的，多数同一氨基酸的密码子均在同一方框内，通常只是第三个碱基不同；在多数情况下，这第三位如果是嘧啶，将编码相同的氨基酸（是同义的），大多数情况下如是嘌呤，其密码子也是同义的。一般来说，如果第二位是嘧啶它所编码的氨基酸是亲水性的，如果第二位是嘌呤则所编码的氨基酸是极性的。

突变的效应

一般认为遗传密码进化方式是以突变影响的最小化进行的。最常见的突变是**转换**，即由一嘧啶突变为另一嘧啶，或者是由一嘌呤转变为另一嘌呤。**颠换**是由嘧啶突变为嘌呤或由嘌呤转变为嘧啶的突变。第三位的转换通常不改变其编码的氨基酸，例外的是会引起蛋氨酸与异亮氨酸、或者色氨酸与终止密码子间的互换。在第三位的颠换一半以上是无效的，其余的通常导致相似氨基酸间的互换，如天冬氨酸或谷氨酸间的互换。在第二位的转换将导致化学性质相似氨基酸间的转变，而颠换将改变氨基酸类型。

第一位上的突变（无论是转换还是颠换）通常编码性质相似的氨基酸，在少数情况甚至是同一氨基酸。

通用性

在遗传密码破译后的相当长一段时间里，大家认为密码子是普遍适用的，即在所有生物中都通用。然而 1980 年后，发现线粒体基因组所使用的密码子与标准或者说“通用”密码稍有不同。其实现在已知一些单细胞生物也使用一些变异了的遗传密码。表 P1.2 列举了这些变异了的遗传密码。

表 P1.2 遗传密码的例外

密码子	通常含义	变异的	细胞器或生物
AGA	Arg	终止密码子，Ser	一些动物的线粒体
AGG			
AUA	Ile	Met	线粒体
CGG	Arg	Trp	植物线粒体
CUN	Leu	Thr	酵母线粒体
AUU	Ile	起始密码子	一些原核生物
GUG	Val		
UUG	Leu		
UAA	终止密码子	Glu	一些原生动物
UAG			
UGA	终止密码子	Trp	线粒体，支原体

可读框

观察 DNA 序列，例如经基因组测序计划获得的序列，无论是用肉眼或是计算机都可辨认出始于 ATG，止于 TGA、TAA 或 TAG 的连续的密码子区域。当没有已知的蛋白产物时，该区域被称为**可读框**，而当确知该可读框编码某一确定蛋白时，它就被称为编码区。也就是说一个可读框是**潜在的编码区**。

重叠基因

尽管通常一个基因编码一条多肽链，而在进化上受到很大限制在一特定区域编码多于一个蛋白有很多，现在知道有许多有重叠编码区（重叠基因）的例子。重叠基因通常是在基因组较小而需要贮存更多信息的情况下产生。例如噬菌体 ΦX 174 基因组只有 5386bp，却编码了 11 种蛋白，其总分子质量达 262kDa。如果基因不重叠，整个基因组至多只能编码 200kDa 的蛋白。该基因组中有三个较短链的蛋白编码区在较长蛋白的编码区内。在原核生物中，核糖体为翻译重叠基因须发现第二个起始密码子，这一步可在不解离模板的状态下实行。而真核生物采用不同的方式起始蛋白合成（参见 Q3），通常采取可变 RNA 加工方式来从一个基因合成出不同蛋白。

P2 tRNA 的结构与功能

要 点

tRNA 的一级结构	tRNA 的线性长度（一级结构）约 60～95 个核苷酸，通常为 76，其中含有许多修饰核苷，特别是胸腺嘧啶核苷、假尿嘧啶核苷、二氢尿苷和肌苷。tRNA 分子中有 15 个不变残基、8 个可变残基。
tRNA 的二级结构	三叶草结构通常是 tRNA 分子二级结构的模型，显示出由于不同区域的碱基配对而形成的四个茎（臂）三个环的结构。5′端与 3′端的碱基配对形成不带环的氨基酸的接受臂。tRNA 的结构按逆时针方向依次为二氢尿嘧啶环（D-loop），反密码子环以及 T 环（T-loop）。大部分恒定和半恒定残基位于环上，而不在茎部。
tRNA 的三级结构	在单链环的碱基间（主要是不变残基）形成九个氢键，将二级结构折叠成 L 形三级结构，其分子的两端分别是反密码子环以及另一端的氨基酸接受臂。
tRNA 功能	tRNA 的 3′端负载特定氨基酸后形成氨酰- tRNA，tRNA 分子在蛋白合成中起转接器作用。
tRNA 的氨酰化	tRNA 分子的负载或氨酰化是通过二步反应完成的。首先氨酰 tRNA 合成酶将腺苷酸（AMP）附着到氨基酸的羟基上，形成氨酰腺苷酸中间体，随后合适的 tRNA 替代 AMP。
氨酰 tRNA 合成酶	合成酶可以是单体、二聚体或四聚体。氨酰 tRNA 合成酶在其 L 形内部与相关 tRNA 发生反应，并利用 tRNA 被称作识别元件的部分鉴别这些相似的 tRNA 分子。
校正	当合成酶完成第一步反应，一个化学结构相似但是错误的氨基酸被氨酰化了时，合成酶不进行第二步反应，而是水解掉氨酰腺苷酸来完成校正。
相关主题	核酸结构（C1）　　tRNA 加工、RNase P 和核酶（O2）

tRNA 的一级结构

tRNA 是将氨基酸转运到核糖体，并破译 mRNA 信息的转接器分子。它们的**一级结构**（即核苷酸的线性序列）长约 60～95 个核苷酸，通常为 76。在任何 tRNA 分子中，均含有许多修饰碱基，有时能占到分子总碱基数的 20%。实际上，在迄今为止已鉴定的几百种 tRNA 分子中，已观察到 50 余种不同类型的修饰碱基，都是在转录后经修饰形成的。图 P2.1 以核苷的形式列出了最常见的七种类型。其中的四种是在 tRNA 分子中常见

的修饰碱基，分别是含有在 RNA 分子中不常见的胸腺嘧啶碱基的**胸腺嘧啶核糖核苷**（T），以及在 RNA 分子中常见的**假尿嘧啶核苷**（ψ)、**二氢尿苷**（D）和**肌苷**（I)。除肌苷外，它们几乎出现在 tRNA 序列中的相似位置。字母 D 和 T 被用来命名二级结构特征（参见下文)。在 tRNA 的一级结构中，有 15 个核苷酸是恒定的，有 8 个位置上的核苷酸是半恒定的，即要么是嘌呤（R）要么是嘧啶（Y)。按标准定位习惯，将第 1 位定为 5′端，而将第 76 位定为 3′端，于是有：

8, 11, 14, 15, 18, 19, 21, 24, 32, 33, 37, 48, 53, 54, 55, 56, 57, 58, 60, 61, 74, 75, 76

D 环（14, 15, 18, 19, 21）　反密码子环（32, 33, 37）　T 环（53, 54, 55, 56, 57, 58, 60, 61）　接受臂（74, 75, 76）

恒定或半恒定核苷酸所处位置在二级结构或三级结构形成中均起作用（参见下文)。

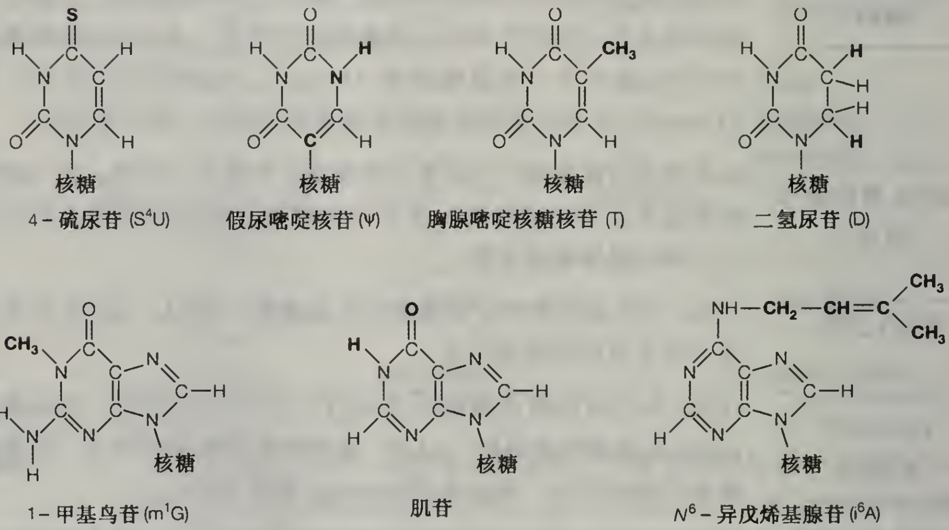


图 P2.1　tRNA 中的修饰核苷

tRNA 的二级结构

所有 tRNA 都有共同的**二级结构**即**三叶草**结构（图 P2.2a 不同区域的碱基配对形成茎环结构)。这一结构含有由 RNaseP 切割形成的 5′-单磷酸而不是通常的 5′-三磷酸（参见 O2)。还有由 tRNA 分子的 3′端与 5′端间的碱基配对而形成的 7bp 茎结构，不过真核生物在加工中添加上去的 74 至 76 位的恒定残基（即末端的 5′- CCA - 3′）并不包含在这一碱基配对区段（参见 O2)。这一茎结构被称为**氨基酸接受臂**。按逆时针方向由 5′向 3′的方向，接下来的二级结构被称为 D 环，由 3 或 4 个碱基对的茎和一个 D 环（二氢尿苷环）组成，该环上常含修饰碱基二氢尿嘧啶。下一个特征结构是由 5 个碱基对的茎和 7 个残基对的环组成，环中含可与mRNA三联体密码序列互补的相邻三核苷酸，即**反密码子**。肌苷在反密码子中的出现使得 tRNA 能够与多于一个的密码序列进行碱基配对（参见 Q1)。接着是**可变臂**，含有 3～21 个残基不等，并且可以形成多至 7 个碱基对的茎状结

构。tRNA 长度变异的另一部位是在图 P2.2a 中以虚线表示的 D 环。二级结构的最后一个主要特征是 **T 臂**或 TΨC 臂，由 5 个碱基对的茎和含有恒定碱基 GTΨC 的环构成。值得一提的是 tRNA 分子中大多数的恒定残基位于环区，在二级结构形成中并不起主要作用，其中的几个有助于三级结构的形成。

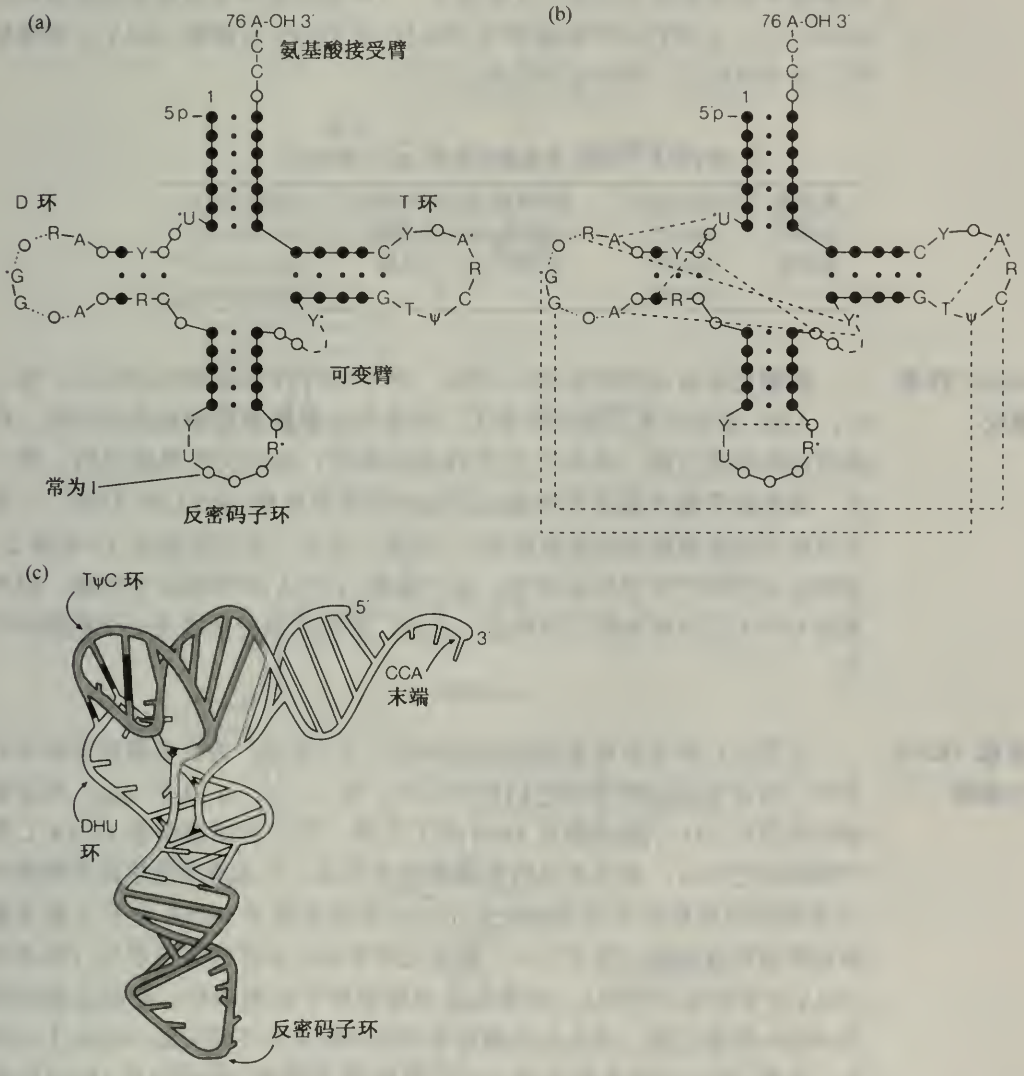


图 P2.2 tRNA 结构。(a) 显示恒定和半恒定核苷酸的三叶草结构，I = 肌苷，ψ = 假尿苷，R = 嘌呤，Y = 嘧啶，* 表示修饰碱基；(b) 用虚线表示的 tRNA 核苷酸间的三级氢键；(c) 酵母 $tRNA^{Tyr}$ 的 L 形三级结构（图 c 引自 D. M. Freifelder (1987) 分子生物学，第二版）。

tRNA 的三级结构

共有 9 个氢键（**三级氢键**）帮助形成 tRNA 分子的**三级结构**。这些氢键主要涉及几个恒定碱基间碱基配对（图 P2.2b）。D 臂上和 T 臂的碱基配对将 tRNA 分子折叠成 L 形，反密码子在 L 形的一端，而氨基酸接受

臂则在另一端。tRNA 三级结构经碱基堆叠作用而得到加强（图 2C）（参见 C2）。

tRNA 功能

经氨酰化反应，tRNA 与氨基酸连接成为**氨酰-tRNA**（负载 tRNA），而蛋白合成中的转接器分子正是这些负载 tRNA。被称为氨酰-tRNA 合成酶特异催化氨酰化反应，且极为专一（即将特定的氨基酸连接到特定的 tRNA 上），这种特定的氨基酸与 tRNA，或 tRNA 与氨酰 tRNA 合成酶被称之为关联组合，其命名列于表 P2.1 中。

表 P2.1 tRNA 合成酶与负载 tRNA 的命名

氨基酸	相关 tRNA	相关氨酰-tRNA 合成酶	氨酰-tRNA
丝氨酸	$tRNA^{Ser}$	丝氨酰-tRNA 合成酶	$Seryl\text{-}tRNA^{Ser}$
亮氨酸	$tRNA^{Leu}$	亮氨酰-tRNA 合成酶	$Leucyl\text{-}tRNA^{Leu}$
	$tRNA^{Leu}_{UUA}$		$Leucyl\text{-}tRNA^{Leu}_{UUA}$

tRNA 的氨酰化

氨酰化反应过程如图 P2.3 所示，是一由 ATP 驱动的两步反应。第一步，AMP 连接到氨基酸的羟基上，形成称为**氨酰腺苷酸**高能中间体，释放出的焦磷酸水解（得到 2 分子的无机磷酸）驱动反应继续进行；第二步，氨酰腺苷酸与**适合**的空载 tRNA 作用形成氨酰 tRNA 和 AMP。一些合成酶可将氨基酸连接到核糖的 2′-羟基，而另一些可连接在 3′-羟基上，连接后这两种产物可相互转换。由于氨酰-tRNA 的键能高于肽键，因此氨酰 tRNA 的形成有助于驱动蛋白合成，因为肽键的形成是一个耗能的反应。

氨酰 tRNA 合成酶

尽管它们都负责将氨基酸连接到相应 tRNA 上，各合成酶却有很大的差异。根据亚基结构可将它们分为四类，即 α、α_2、α_4 以及 $\alpha_2\beta_2$。其多肽链的长度从 334 个氨基酸至 1000 多个不等。这些合成酶是在 tRNA L 形的侧面与之结合，并有各自的氨基酸结合位点。合成酶须能够区分细胞中 40 多种形状相似的不同 tRNA 分子，这主要依赖于 tRNA 分子上称为**鉴别元件**的特殊部位（图 P2.4）。鉴别元件并非总是反密码子序列（在不同 tRNA 分子中是不同的），常常还包含接受臂中的碱基对。如果这些元件在 tRNA 间被互换，那么合成酶就会将氨基酸接在不对应的 tRNA 上。例如，如果 $tRNA^{Ala}$的鉴别元件 G3∶U70 被用来替换 $tRNA^{Cys}$或 $tRNA^{Phe}$的 3∶70碱基对，那么这些被修饰过的 tRNA 就会被丙氨酰-tRNA 合成酶所识别，结果负载丙氨酸。

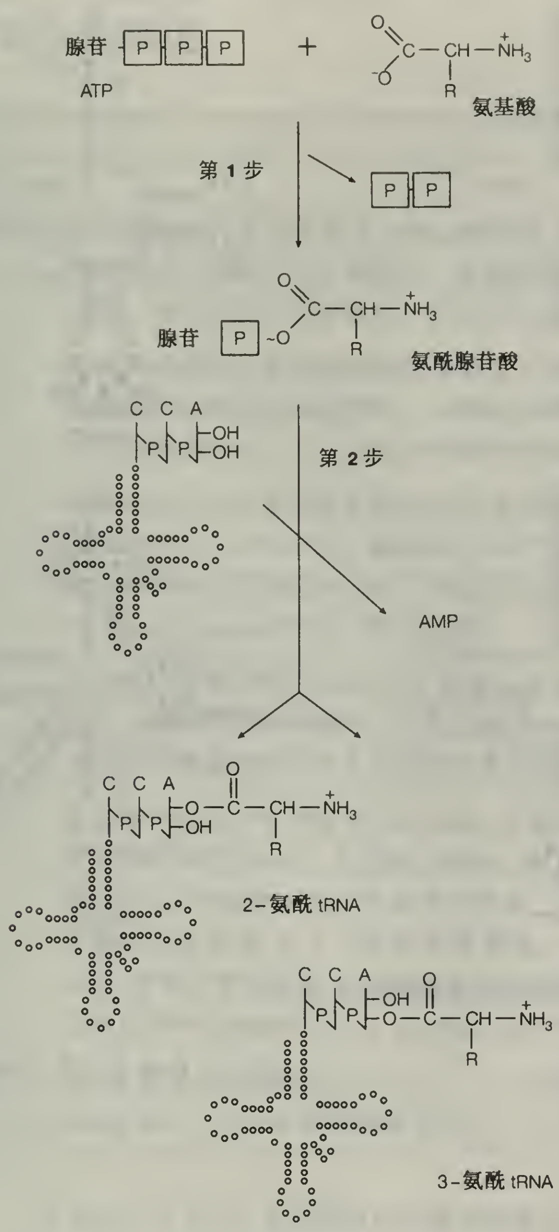


图 P2.3　氨酰-tRNA 的形成

校正

某些须对两种化学结构相似的氨基酸进行区分的合成酶行使校正功能。如果偶然在进行第一步反应时将错误的氨基酸氨酰化，那么它们将不再进行第二步反应，而是进行氨酰腺苷酸水解。只有当单一识别过程不足以区分不同氨基酸时，校正才是必需的。苯丙氨酸和酪氨酸在第一步即可被区分开来，因为苯环上羟基位置的不同，因此这种情况下就不需校正。

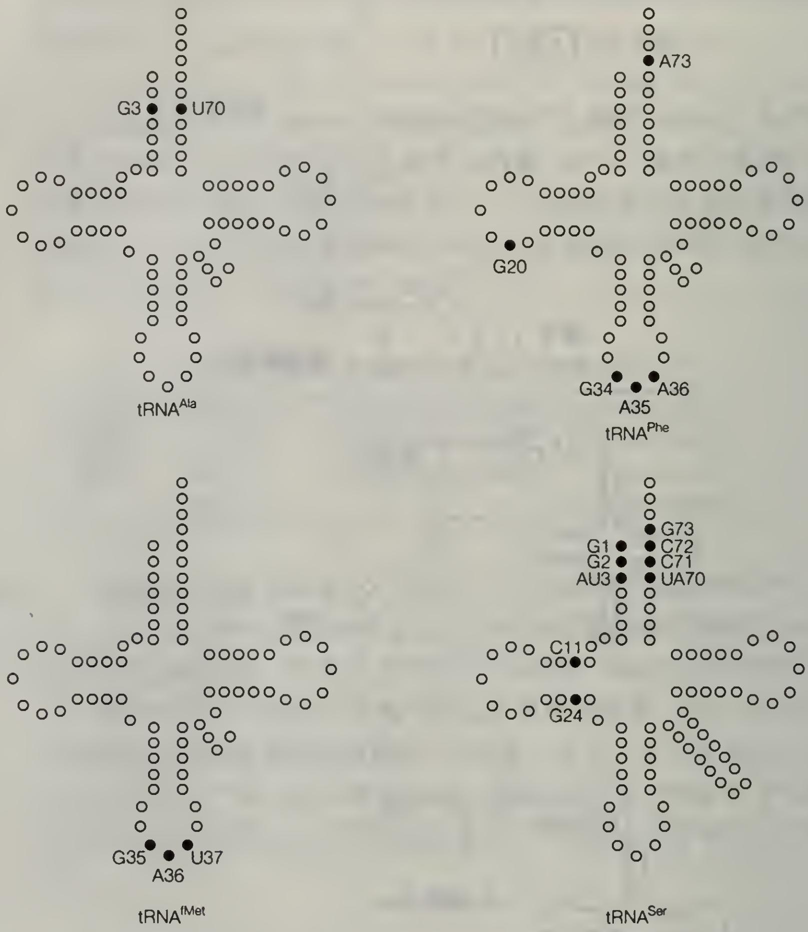


图 P2.4 不同 tRNA 分子中的鉴别元件

（林会兰 译 刘进元 校）

Q1 蛋白质合成概述

要 点

密码子与反密码子的相互作用	在核糖体的大小亚基的缔合处，mRNA 的密码子与 tRNA 上的反密码子平行形成 3 个碱基对。如果反密码子的 5′碱基是被修饰过的，则其 tRNA 通常能与一个以上的密码子相互反应。
摆动现象	摆动假说描述了被修饰过的反密码子 5′碱基能与密码子的 3′碱基形成非标准型的碱基配对。当摆动位核苷是肌苷时，tRNA 能与 3′端碱基为 A、C、U 的 3 种密码子形成碱基配对。
核糖体结合位点	核糖体结合位点是原核生物 mRNA 起始密码子上游的一段序列，它能够与 16S rRNA 的 3′端附近的互补序列配对，对核糖体进行定位以便起始蛋白质的合成。若以其发现者的名字来命名则称之为 Shine-Dalgarno 序列，即 SD 序列。
多聚核糖体	当核糖体连续不断地向 mRNA 附着，并进行翻译且沿 mRNA 移动时，就形成多聚核糖体。多聚核糖体是位于 mRNA 分子链上处于不同翻译阶段的多个核糖体的复合体。
起始 tRNA	在原核和真核生物中能够识别 AUG 起始密码子，起动蛋白合成的一种特殊的 tRNA。在原核生物中，起始 tRNA 首先在甲硫氨酰 tRNA 合成酶的作用下负载甲硫氨酸，接着甲硫氨酸残基经转甲酰酶作用转变为 *N*－甲酰甲硫氨酸。在真核生物中，起始 tRNA携带的甲硫氨酸未经修饰。在大肠杆菌中，起始 tRNA 与运送内部甲硫氨酸的 tRNA 在结构上是不同的。
相关主题	rRNA 加工与核糖体（O1）　tRNA 的结构与功能（P2） tRNA 加工、RNase P 和核酶（O2）　蛋白质合成机制（Q2）

密码子与反密码子的相互作用

当 mRNA 与 tRNA 在核糖体的凹槽处相遇时，tRNA 一端的**反密码子**将与 mRNA 上的互补的三联体碱基即**密码子**相互作用（参见 O1）。这种作用具反向平行的特征（图 Q1.1）。研究发现某些高纯度的 tRNA 能与多于一个的密码子作用，这与反密码子 5′位存在修饰核苷尤其是肌苷相关（参见 P2）。5′位的肌苷是经腺苷的转录后加工（参见 O2）而形成，该反应由可将 6－氨基转变为酮基的**反密码子脱氨酶**催化完成的。

摆动现象

为解释遗传密码的丰余性，克里克提出了摆动假说。通过模型分析，克里克认识到反密码子的 5′碱基相对于另外两个碱基有更大的活动范围，

因此只要核糖间的距离接近正常距离，它能够形成**非标准碱基对**。他的划时代预言及其实际观察结果列于表 Q1.1。嘌呤-嘌呤或嘧啶-嘧啶配对是不允许的，因为核糖之间的距离不合适，没有一种 tRNA 能够识别三个以上的密码子，因此，至少需要 32 种 tRNA 去破译 61 个密码子（终止密码子除外）。tRNA 依其摆动碱基（反密码子的 5′碱基）能够识别一个、两个或三个密码子。如果 5′碱基是 C，tRNA 只能识别以 G 结尾的密码子；如果是 G，则可将识别以 U 和 C 结尾的两种密码子；如果是 U，它将随即被修饰，可与 A 或 G 配对。摆动核苷决不会是 A，因为通常 A 被转成能与 A、C 或 U 配对的肌苷。

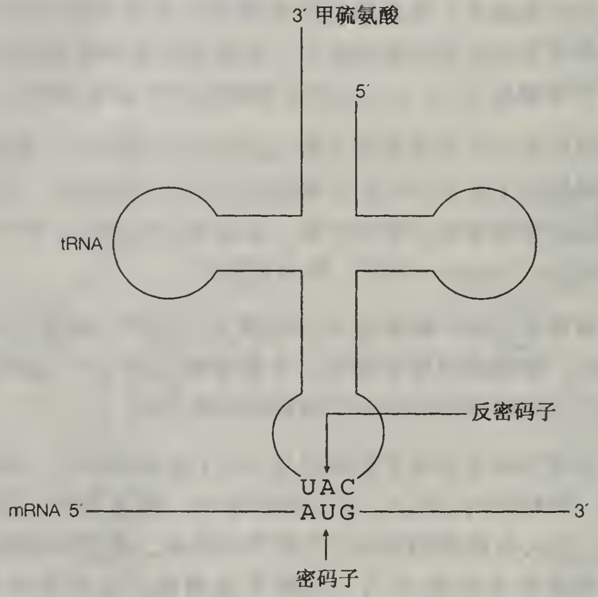


图 Q1.1　密码子与的反密码子的相互作用

表 Q1.1　最初的摆动预测

反密码子的 5′碱基	预测的 3′密码子碱基	实际观察
A	U	经反密码子脱氨酶将 A 修饰为 I
C	G	没有摆动，正常碱基配对
G	C 和 U	G 及修饰 G，能与 C 和 U 配对
U	A 和 G	未发现反密码子 5′碱基是 U
I	A、C 和 U	按所预言的规律摆动，I 能识别 3′- A，- C 或- U

核糖体结合位点

在原核生物 mRNA 起始密码子上游 8～13 个核苷酸处有一保守序列。该序列由 Shine 和 Dalgarno 发现，是一通常含有 5′- AGGAGGU - 3′序列的全部或部分的多嘌呤序列。实验证实这段序列可与核糖体小亚基中的 16S rRNA 的 3′端序列 5′- ACCUCCU - 3′进行碱基配对，因此称之为**核糖体结合位点**或 Shine-Dalgarno 序列即 **SD 序列**。SD 序列的作用被认为是相

对于起始密码子而正确定位核糖体。

多聚核糖体　当核糖体开始对 mRNA 分子进行翻译（参见 Q2），并从起始密码子处移动约 70～80 个核苷酸后，第二个核糖体能够在核糖体结合位点处组装并开始翻译；而当第二个核糖体沿 mRNA 移动，第三个核糖体又开始组装，如此往返，结果在 mRNA 上聚集多个核糖体。特将单一 mRNA 分子上的多个核糖体称为**多聚核糖体**。某些 mRNA 分子上会聚集有 50 个核糖体之多，尽管它们之间相隔约 80 个核苷酸。

起始 tRNA　研究发现，无论在原核生物中还是真核生物中，掺入蛋白质多肽链的第一个氨基酸都是甲硫氨酸，只不过在原核生物中甲硫氨酸被修饰为 *N*-**甲酰甲硫氨酸**。在两类生物中，AUG 起始密码子均由一个特定**起始 tRNA** 所识别。起始 tRNA 与识别编码区内其他 AUG 密码子的 tRNA 有所不同，如在原核生物大肠杆菌中，这两种 tRNA 之间有微妙的差别（图 Q1.2）。起始 tRNA 的碱基配对有更大的灵活性，由于它的反密码子环上缺少烷基化的 A，因此可识别 AUG 和 GUG 起始密码子，而 GUG 是原核生物的 mRNA 中很少出现。而非起始 tRNA 缺乏灵活性，只能与 AUG 配对。这两种 tRNA 均由同一**甲硫氨酰-tRNA 合成酶**催化完成，甲硫氨酸的负载而形成甲硫氨酰-tRNA，但只有起始甲硫氨酰 tRNA 被转**甲酰酶**修饰，而转变为 *N*-甲酰甲硫氨酰-tRNA。*N*-甲硫基与肽键相似，会有助于起

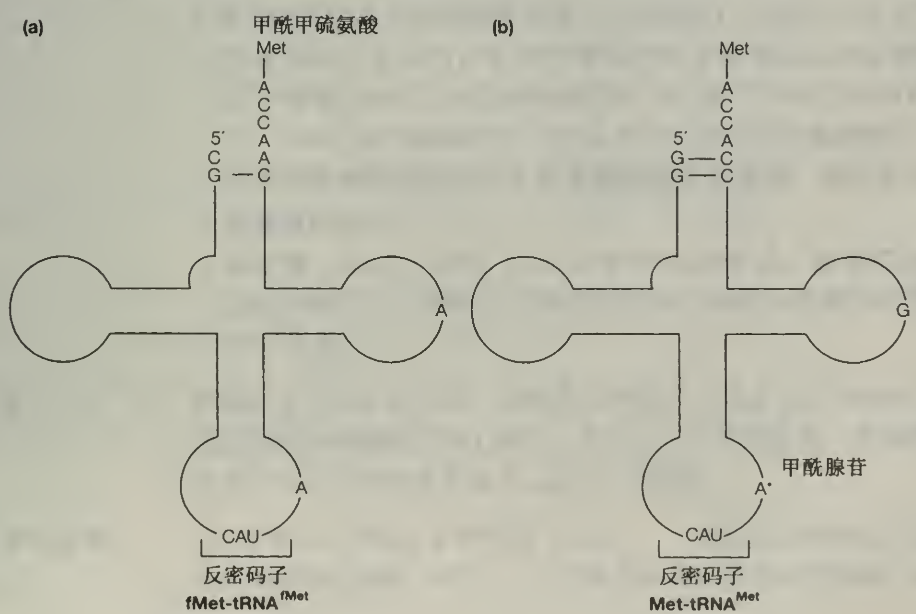


图 Q1.2　*E. coli* 的甲硫氨酰 tRNA。（a）起始 tRNA，$fMet\text{-}tRNA^{fMet}$；（b）甲硫氨酰-tRNA，$Met\text{-}tRNA^{Met}$

始 tRNA 进入核糖体的 P 位点，而其他 tRNA 则进入 A 位点（参见 Q2）。

Q2 蛋白质合成机制

要 点

基本过程 蛋白质的合成过程可分为三步：

- **起始** 核糖体在 mRNA 上的组装；
- **延伸** 氨基酸的转运、肽键形成及沿 mRNA 移动（移位）的重复循环；
- **终止** 多肽链的释放。

起 始 在原核生物中，蛋白质合成的起始需要核糖体大小亚基、mRNA、起始 tRNA、3 种起始因子（IF）及 GTP。IF_1 和 IF_3 与 30S 亚基结合，以阻止大亚基的结合；接着 IF_2 和 GTP 与小亚基结合，以利于随后的起始 tRNA 的结合；形成的小亚基复合体方能经由核糖体结合点附着到 RNA 上。而后起始 tRNA 与 AUG 起始密码子配对并释放 IF_3，最终形成 30S 起始复合体。接着大亚基与该复合体结合，替换 IF_1 和 IF_2 + GDP，形成 70S 起始复合物。这是在 mRNA 正确部位组装的完整核糖体。

延 伸 延伸涉及三个延伸因子（EF：EF-Tu、EF-Ts 和 EF-G）、GTP、负载 tRNA 以及 70S 起始复合体（或等同物）。反应分三步进行：

- 负载 tRNA 与 EF-Tu 和 GTP 形成的复合体被运送至核糖体，GTP 水解、EF-Tu·GDP 释放出来。在 EF-Ts 和 GTP 的作用下，Tu·GDP 可被再利用（经由 EF-Tu-EF-Ts 交换循环）。
- 肽酰转移酶将相邻的两个氨基酸相连形成肽键，该过程不需要能量的输入。
- 移位酶（EF-G）利用 GTP 水解释放的能量，使核糖体沿 mRNA 移动一个密码子，释放出空载 tRNA 并将新生肽链运至 P 位点。

终 止 释放因子（RF1 或 RF2）识别终止密码子，并在 RF3 的作用下，促使肽酰转移酶在肽链上加上一个水分子并释放肽链。核糖体释放因子有助于核糖体亚基从 mRNA 上解离。

相关主题 tRNA 加工、RNase P 和核酶（O2）　蛋白质合成概述（Q1）
tRNA 结构与功能（P2）　真核生物蛋白质合成的起始（Q3）

基本过程 蛋白质合成过程可分为三步：

- **起始**——核糖体在 mRNA 分子上的组装；

- **延伸**——氨基酸添加的重复循环；
- **终止**——新生蛋白（多肽）链的释放。

如图 Q2.1～Q2.3 所示该过程还涉及一系列因子。在原核生物中，**起始和延伸因子**分别缩写成 IF 及 EF，而在真核生物中则分别称为 eIF 和 eEF。原核生物与真核生物的蛋白质合成机制在细节上有明显的不同，这主要表现在起始阶段。鉴于此，本节将讨论原核生物中的合成机制，下一节（Q3）将详细描述真核生物的蛋白质合成过程。

起始

起始阶段的主要任务是在 mRNA 分子的正确起始点即**起始密码子**处完成完整核糖体的组装。涉及的组成元件有**大小核糖体亚基**、**mRNA**、已负载的**起始 tRNA**、**三个起始因子**及 **GTP**。起始因子 IF_1、IF_2 及 IF_3 在数目上均是核糖体的十分之一，其分子质量分别为 9、120 及 22kDa，只有 IF_2 与 GTP 结合。尽管详细过程有待深入研究，但整个过程如下（图 Q2.1）：

- IF_1 和 IF_3 与游离的 30S 亚基结合，以阻止在与 mRNA 结合前 30S 亚基与大亚基的结合，从而防止无活性核糖体的形成。
- 接着 IF_2 与 GTP 的复合体结合到小亚基上，这将有助于已负载的起始 tRNA 与该复合体结合。
- 30S 亚基利用 mRNA 分子上的核糖体结合位点（RBS）附着到 mRNA 上（参见 Q1）。
- 接着起始 tRNA 通过其反密码子与 mRNA 分子上 AUG 密码子的碱基配对，与上述复合体结合，同时释放出 IF_3。IF_3 的作用在于保持大小亚基彼此分离状态，以及有助于 mRNA 结合。此时的复合体称为 **30S 起始复合体**。
- 此时 50S 亚基将与上述复合体结合，替换出 IF_1 和 IF_2，而 GTP 在此耗能过程中被水解。起始后期形成的该复合体被称为叫 **70S 起始复合体**。

如图 Q2.1～Q2.3 所示，组装完的核糖体有两个 tRNA 结合位点，分别称为 **A 位点**和 **P 位点**，即氨酰基位点和肽酰基位点。A 位点是氨酰-tRNA 结合的位点，而 P 位点则是延伸肽链所在处。两个位点均位于小亚基的凹槽处（参见主题 O1），含有正在被翻译的相邻密码子。起始的主要成果是将起始 tRNA 置于 P 位点。只有起始 tRNA 能进入该位点，其他 tRNA 必须进入 A 位点。

延伸

随着 70S 起始复合体的形成，**延伸循环**随即开始。延伸过程也可细分为三步：①**氨酰 tRNA 的转运**；②**肽键的形成**；③**移位**。如图 Q2.1 所示，延伸时 P 位点负载有肽酰-tRNA，而 A 位点处于空载状态。延伸过程涉及 3 个**延伸因子**即 **EF-Tn**、**EF-Ts** 和 **EF-G**，它们均能与 GTP 或 GDP 结合，分子质量依次为 45、30 和 80kDa。EF-Ts 和 EF-G 在数量上与核糖

体相当，而 EF-Tn 约为核糖体数的 10 倍。

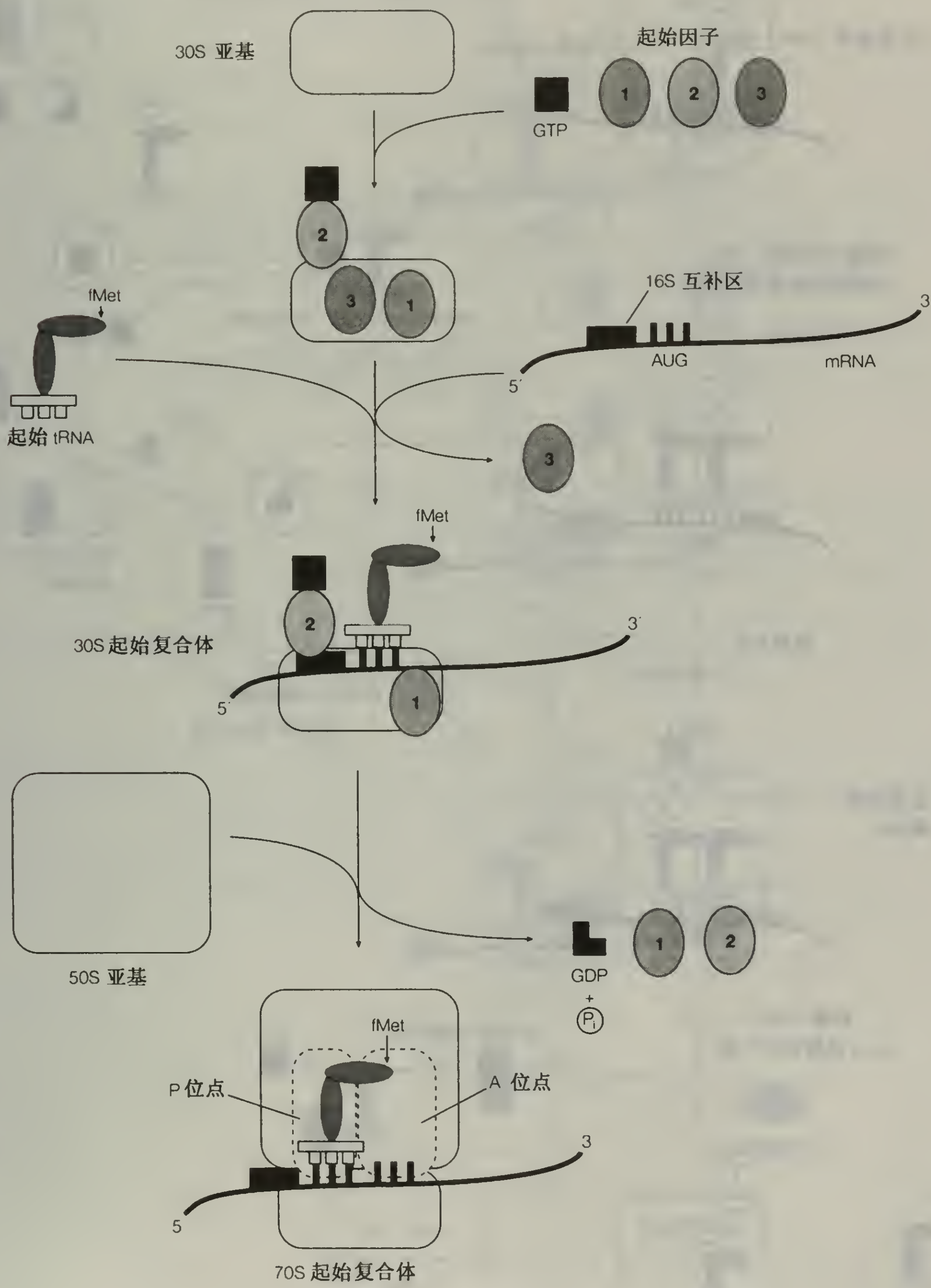


图 Q2.1 *E. coli* 中蛋白质合成的起始

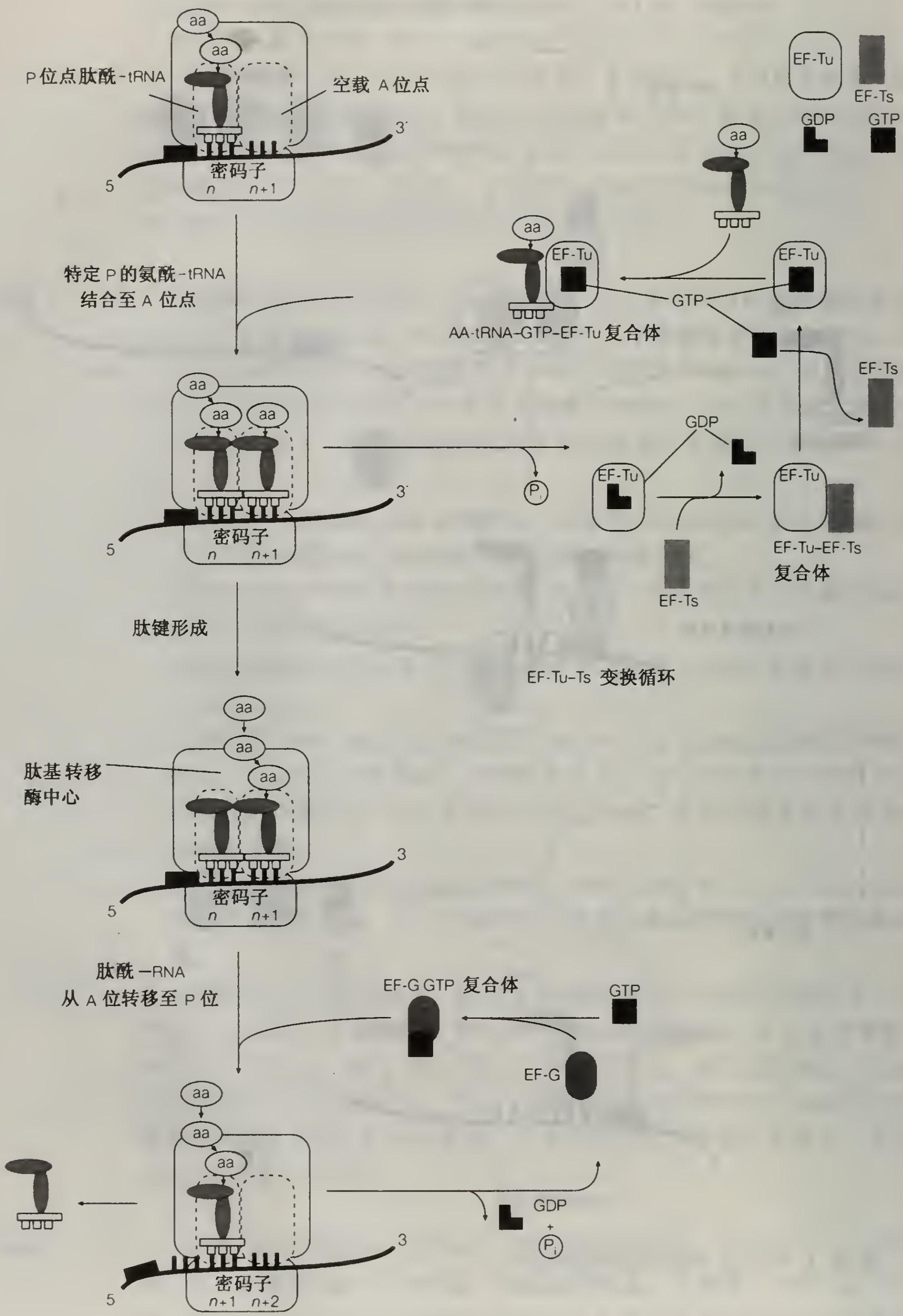


图 Q2.2 蛋白质合成的延伸期

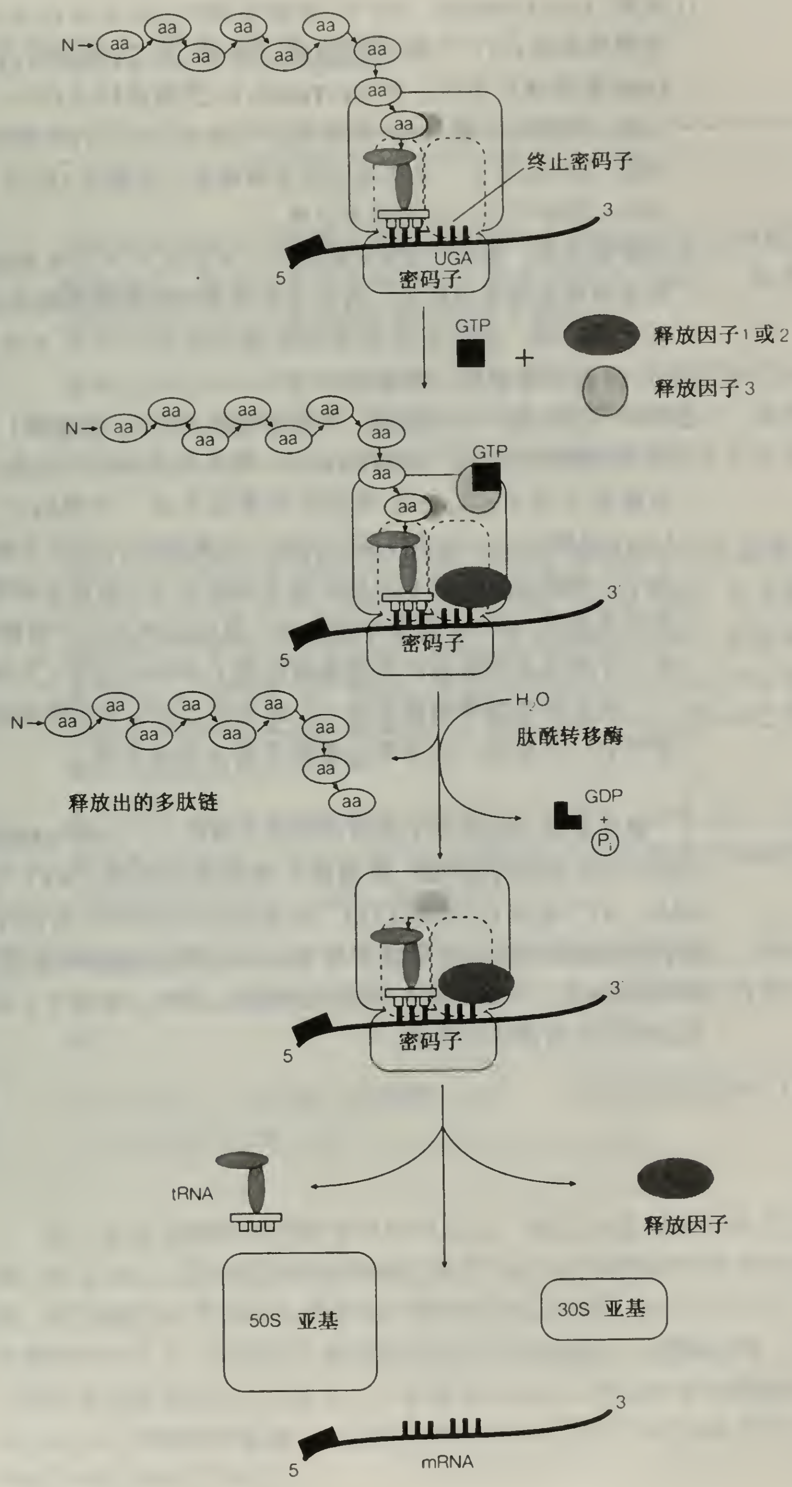


图 Q2.3　*E. coli* 中蛋白质合成的终止

①氨酰-tRNA的转运。EF-Tn为氨酰tRNA进入A位点所必须的，此过程利用水解GTP所释放的能量。在EF-Ts的作用下，释放的**EF-Tn·GDP复合体**可再生。在**EF-Tn-EF-Ts交换循环**途径中，EF-Ts替换了GDP，但随后又被GTP所取代。产物EF-Tn·GTP复合体可与另一个氨酰-tRNA结合，并将其运送至核糖体。除起始tRNA外，所有氨酰-tRNA都能与EF-Tu形成复合体。

②肽键的形成。氨酰-tRNA进入后，A位点和P位点都满载，两个要相连接的氨基酸靠得很近，此时50S亚基的**肽酰转移酶**可在两个氨基酸之间形成肽键，该反应无需任何能量的补给，因为ATP的能量在活化tRNA时已被储存（参见P2）。

③移位。移位是一耗能过程。具体过程是EF-G（**移位酶**）和GTP的复合体与核糖体结合后，卸载的tRNA从P位点解离，肽酰-tRNA从A位点移至P位点，mRNA相对于核糖体移动一个密码子。最后GDP和EF-G被释放，而后者将被再利用。在腾空的A位点上将出现一个新密码子。最新证据表明在原核生物中卸载的tRNA首先移到**E位（去位）**，然后当氨酰-tRNA结合时而脱离，通过这种方式，核糖体与mRNA通过6个碱基对保持结合而很好地降低了移码的概率（见R4）。

综上所述延伸循环中的一个循环的三步反应至此全部完成，这样的循环会一直重复，直至终止密码子在A位点出现。

终止

通常没有tRNA分子能够识别终止密码子，而称为释放因子的蛋白因子可与终止密码子作用，致使新生多肽链的释放。RF1可识别UAA和UAG，RF2识别UAA和UGA，而RF3有助于RF1或RF2的活性。释放因子使肽酰转移酶将多肽链转至H_2O分子而不是通常的氨酰-tRNA、释放出mRNA、并与核糖体亚基完全解离。此时IF_3可与小亚基结合以阻止无活性70S核糖体的形成。

Q3 真核生物蛋白质合成的起始

要 点	
基本过程	原核生物与真核生物蛋白质合成机制的主要差别在于起始阶段。除此以外，那就是真核生物只有一种释放因子（eRF），真核生物的起始 tRNA 不像原核生物那样被甲酰化。
扫　描	真核生物的 40S 核糖体亚基复合体与 mRNA 的 5′帽结构区域结合，并沿着 mRNA 移动（扫描）寻找 AUG 起始密码子。起始点并不一定总是第一个 AUG 密码子，因为其周围还必须有适当的序列。
起　始	原核生物与真核生物蛋白质合成过程的主要差别在于起始阶段，至少有 9 种 eIF 相关。按功能可将这些因子分为几组，即与核糖体亚基结合的或与 mRNA 结合的、运送起始 tRNA、或是其他替换因子。与原核生物相比，真核生物的起始在与 mRNA 结合前起始 tRNA 先与 40S 亚基结合。转运起始 tRNA 的 eIF2 的磷酸化是重要的被调控对象。
延　伸	蛋白质合成在延伸阶段基本上与原核生物相同（参见 Q2）。与原核生物的 EF-Tu、EF-Ts 和 EF-G 具相同功能的真核生物延伸因子分别称为 eEF1α、eEFβγ 和 eEF2。
终　止	在真核生物的终止过程中只使用一种释放因子（eRF），该因子需要 GTP 以终止蛋白合成，并能识别三种终止密码子的任何一种。
相关主题	tRNA 加工、RNA 酶 P 和核酶（O2）　蛋白质合成机制（Q2） tRNA 结构与功能（P2）

基本过程　除了在真核细胞的线粒体和叶绿体内（被认为是起源于共生原核生物，参见 A2），蛋白质合成的详细机制也与在 Q2 所描述的原核生物的不同，其主要区别是在涉及**许多 eIF 因子**的起始阶段。由于 mRNA 上没有核糖体结合位点，要找到正确的起始密码子需要经过一**扫描过程**。有两种不同的甲硫氨酸 tRNA，其中之一是起始 tRNA，结合的甲硫氨酸并不被转变成 ***N*-甲酰甲硫氨酸**。在原核生物和真核生物蛋白质合成中所涉及各种因子的比较列于表 Q3.1。

表 Q3.1 原核真核生物蛋白质合成因子的比较

原核的	真核的	功 能
起始因子		
IF1、IF3	eIF3、eIF4C、eIF6	与核糖体亚基结合
	eIF4B、eIF4F	与 mRNA 结合
IF2	eIF2、eIF2B	起始 tRNA 传递
	eIF5	替换其他因子
延伸因子		
EF-Tu	eEF1α	传递氨酰 tRNA 至核糖体
EF-Ts	eEF1βγ	EF-Tu 或 eEF1α 的再循环
EF-G	eEF2	位移
终止因子		
RF1		
RF2	eRF	多肽链的释放
RF3		

扫描

由于真核生物 mRNA 上没有 Shine-Dalgarno 序列（SD 序列），所以选择起始密码子的机制当然不同。Kozak 提出了**扫描假说**，认为已与起始 tRNA 结合的 40S 亚基，附着到 mRNA 的 5′端，并沿着 mRNA 移动，直至找到合适的 AUG 起始密码子。合适的 AUG 并不一定总是第一个 AUG，因为还需要有正确的**上下游序列**（5′- CCR CCAUGG - 3′），这里 R 代表嘌呤。

起始

图 Q3.1 显示了在真核生物蛋白质合成的起始阶段所涉及到的步骤及因子。在真核生物蛋白质合成中至少涉及有九种已被鉴定的起始因子，其中一些与三种原核生物的起始因子有相似功能。根据它们的功能，可将其分组如下：

- **与核糖体亚基结合**的因子，如 eIF6、eIF3 和 eIF4C；
- 与 **mRNA 结合**以识别 5′帽子结构并解开二级结构的，如 eIF4B 和 eIF4F。eIF4F 是帽子结合蛋白即 eIF4E 和 eIF4A 的复合体；
- 与**起始 tRNA 输送**相关的，如 eIF2 和 eIF2B；
- **其他可替换因子**，如 eIF5，该因子能替代另外两种因子使 60S 亚基得以结合。

从游离的 40S 亚基和具 5′帽子结构的 mRNA 分子开始的蛋白质合成过程如下：**eIF3** 和 **4C** 结合到 40S 亚基上，促使其与**三元复合体（起始 tRNA、eIF2 和 GTP）**结合。请注意真核生物中的不同装配顺序，在 mRNA结合之前起始 tRNA 先与小亚基结合（与 Q2 相比较，图 Q3.1）。在上述复合体与 mRNA 结合之前，mRNA 先要与 **eIF4B** 和 **4F**（4F 具有识别 5′帽子结构的功能）作用，利用来自 ATP 的能量解旋去掉二级结构。在 40S 亚基复合体通过 5′帽子结构与 mRNA 复合体结合后，利用 ATP 能量开始扫描 mRNA 以寻找 AUG 起始密码子。起始点通常是第一个 AUG。

eIF2 和 eIF3 被 **eIF5** 替换后 **60S** 亚基方能与 40S 亚基结合，并水解 GTP。eIF4C 在帮助 60S 亚基结合形成完整的 **80S 起始复合体**后即被释放。

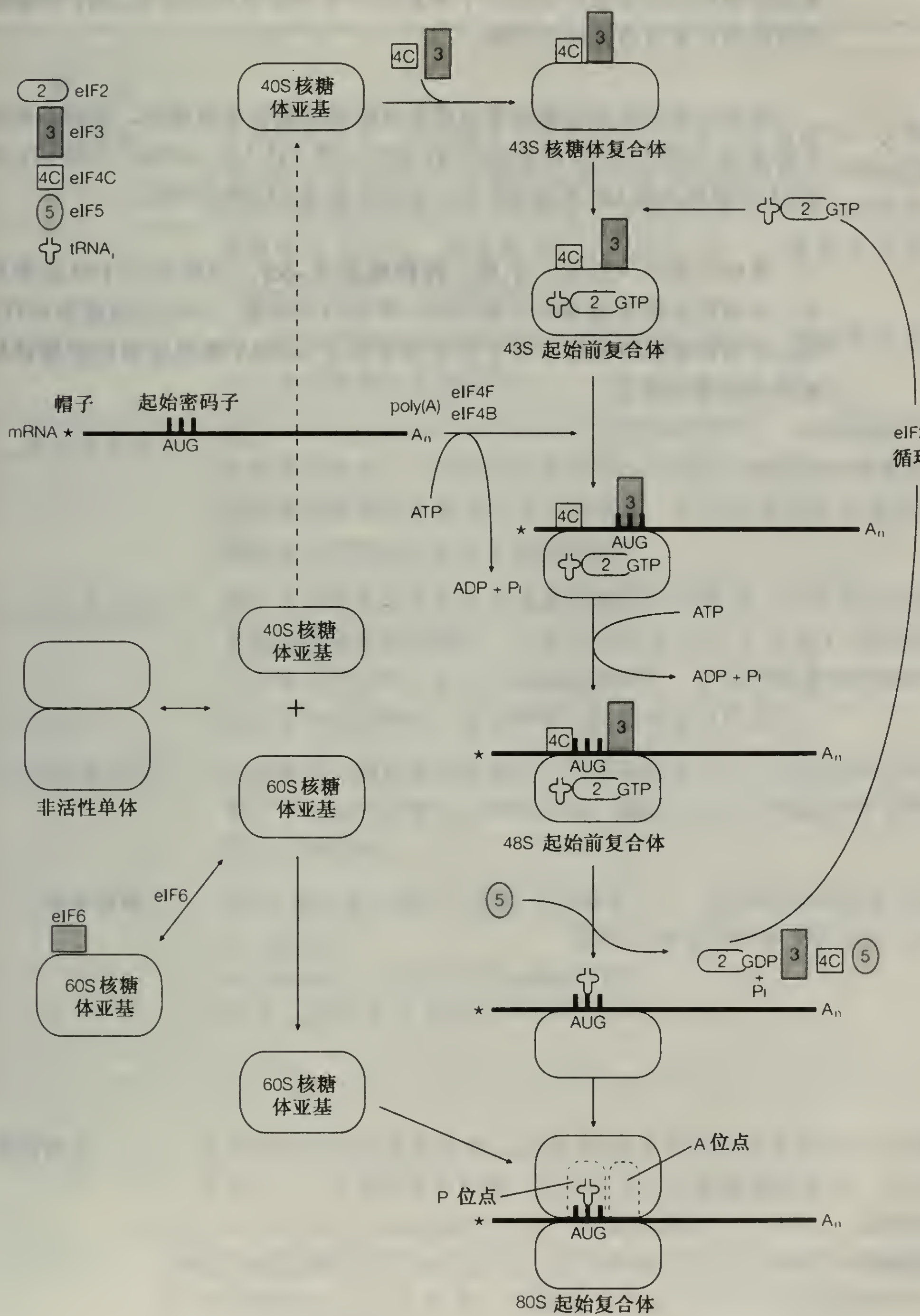


图 Q3.1 真核生物蛋白质合成的起始

释放后的 eIF2 · GDP 复合体在 **eIF2B** 的作用下进入下一轮。循环的速率（也即蛋白质合成的起始速率）被 eIF2 的 α 亚基的磷酸化所调控。某些生物例如病毒感染及随后干扰素的产生等反应均可促进 **eIF2 的磷酸化**而导致对蛋白质合成的抑制。

延伸

原核生物与真核生物的蛋白质合成的延伸循环非常相似，所需三种因子的性质与原核生物相应因子的相似（表 Q3.1）。eEF1α、eEF1βγ 和 eEF2 分别具有在 Q2 所述 EF-Tu、EF-Ts 和 EF-G 相应功能。

终止

真核生物中，仅有一个**单一的释放因子 eRF**，可识别三个终止密码子，并执行原核生物 RF1（或 RF2）和 RF3 的功能。eRF 活性需要 GTP，但还不清楚在真核生物中是否存在为亚基与 mRNA 解离所需的核糖体释放因子的等效因子。

Q4 翻译调控与翻译后加工

要 点

翻译调控 在原核生物中，不同顺反子的翻译水平受下列因素影响：①较短反义分子的结合；②多顺反子 mRNA 对核酸酶的相对稳定性；③阻碍核糖体附着的蛋白质结合。在真核生物中，蛋白质结合也能够掩盖 mRNA，阻碍翻译。5′- AUUUA - 3′重复序列会使 mRNA 不稳定，降低翻译效率。

多蛋白 单一翻译产物被切割形成两个或多个独立蛋白，则被称为多蛋白。许多病毒产生多蛋白。

蛋白质定位 蛋白质中某些短肽序列可决定蛋白质的细胞定位，如细胞核、线粒体或叶绿体。分泌蛋白的信号序列导致执行翻译的核糖体某些使得核糖体停泊在膜上的因子的结合，并将合成蛋白转至膜外。通常信号序列随后被信号肽酶切掉。

蛋白质修饰 新生多肽的最常见的加工是被切割及化学修饰。信号肽的切除、多蛋白成熟片段的释放、中间肽的切除以及 N 端和 C 端的修剪均需要进行切割。除 6 种氨基酸侧链外，所有的氨基酸侧链都可发生多种化学修饰。通常磷酸化调控着蛋白质活性。

蛋白质降解 对损害的、被修饰的或遗传上不稳定的蛋白，可用共价结合的泛素分子标记来观察蛋白降解过程。随后 26S 蛋白酶复合体会将遍在蛋白质降解。

相关主题

RNA 聚合酶Ⅱ基因：启动子与增强子（M4）
mRNA 加工、hnRNP 和 snRNP（O3）
可变 mRNA 加工（O4）
蛋白质合成机制（Q2）
真核生物蛋白质合成的起始（Q3）

翻译调控

由于原核生物与真核生物 mRNA 的性质不同（即多顺反子与单顺反子，参见 L1），且前者没有核膜，所以有不同的**翻译调控方式**。在原核生物中，mRNA 某一区段形成的结构可掩盖核糖体结合位点，这样相对其他顺反子而言就降低了翻译速率。茎环结构的形成可**抑制外切核酸酶**攻击，使多顺反子 mRNA 的某一区域相对其他区域而言有更长的半衰期（因而有更多的翻译机会）。在 *E. coli* 中，几个编码核糖体蛋白的操纵子显示出一种有趣的翻译调控方式，即 mRNA 的某一区域能形成类似于由

该 mRNA 所编码核糖体蛋白的结合位点的三级结构，如果没有足够的 rRNA 供翻译产物结合，这些产物将结合到自身的 mRNA 上，并阻碍其进一步翻译。原核生物有时会产生较短的、可与某些 mRNA 上的核糖体结合位点附近的序列形成双链的**反义 RNA** 以抑制翻译。

真核生物通常用改变基因的转录水平（参见 M4）和（或）RNA 的加工（参见 O3 和 O4）来控制特定蛋白质的合成量，但也有一些调控发生在细胞质中。通常在 3′非编码区存在多个 **5′-AUUUA-3′**单元，是 mRNA 快速降解的标志，因而限制翻译。另一种形式的翻译调控是蛋白质直接结合到 mRNA 上并阻止翻译。这种 RNA 被称为"**掩蔽 mRNA**"，在适当的条件下，当蛋白解离时 mRNA 即可被翻译。一些非编码序列可使得 mRNA定位在细胞质特定部位，当翻译时，能产生细胞内的蛋白质浓度梯度。

多蛋白

噬菌体、病毒转录物（参见 R2）以及许多真核生物编码激素的 mRNA（如阿里皮素原）翻译后先生成一条多肽链，随后被特定的蛋白酶切割，结果从一条翻译产物生成多个成熟蛋白质。原来的多肽链被称为**多蛋白**。

蛋白质定位

研究发现蛋白在细胞内的最终定位取决于其自身蛋白的特性、相对长短以及其氨基酸序列。这些序列可决定蛋白是被分泌、转运至核内还是其他细胞器内。真核细胞的结构越复杂（参见 A1）就意味着有越多的定位类型。无论原核生物还是真核生物，**蛋白分泌**涉及新生蛋白的**信号序列**以及特殊蛋白，包括能识别信号序列的一种核糖核蛋白颗粒即**信号识别颗粒**（SRP）（图 Q4.1）。

如果胞内核糖体开始翻译编码分泌蛋白的 mRNA，SRP 会与核糖体和新生的多肽链结合以停止翻译。SRP 能够识别含有**信号肽**序列的新生肽链的核糖体。信号肽一般由 13 至 36 个氨基酸构成，至少有一个带正电的残基，其后是约 10 至 15 个残基的疏水核，接着为小的中性残基，通常为 Ala。**SRP** 和**受阻的核糖体**一起结合到**内质网**外侧的受体（**SRP 受体**或**停靠蛋白**）上（参见 A2），当核糖体附着到内质网上的**核糖体受体蛋白**后，SRP 即被释放并可被循环利用。核糖体继续翻译，新生肽链被运送到内质网的腔内。在穿过内质网时，信号肽被信号肽酶切除。蛋白释放到内质网腔内后，通常被修饰，通常是**糖基化**修饰。不同的糖基化修饰似乎控制着蛋白质的最终定位。

蛋白质中的其他序列也决定蛋白质的胞内定位，如不同的 N 端序列能导致蛋白质进入线粒体或叶绿体；内部序列-Lys-Lys-Lys-Arg-Lys，或任何五个连续的带正电荷氨基酸，是**核内定位信号**（NLS），能导致含该序列的蛋白（如组蛋白）被运进核内。

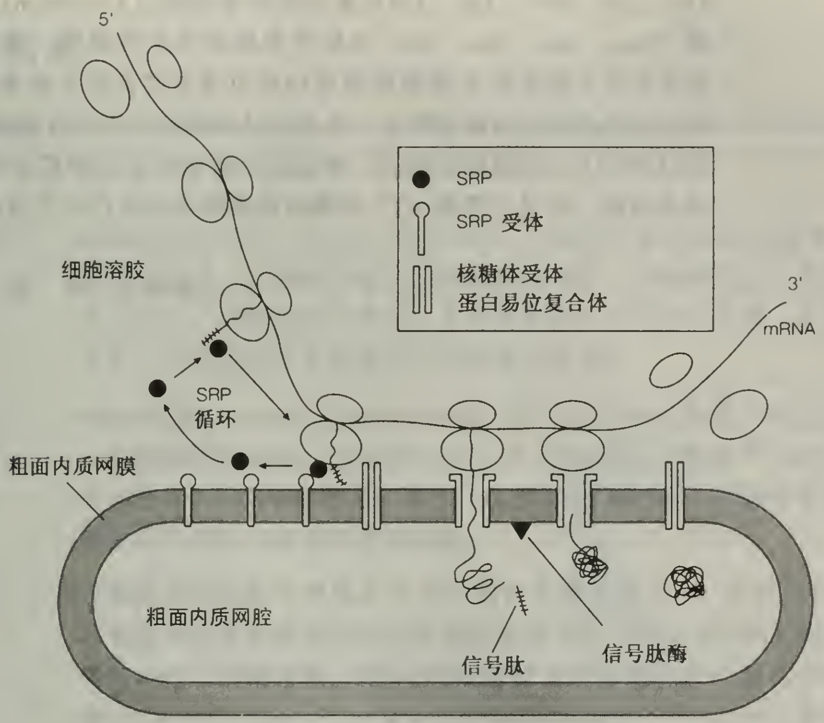


图 Q4.1 真核生物中的蛋白质分泌

蛋白质修饰

新生多肽未必立即成为有功能的蛋白（参见 B2），除了正确折叠及可能形成的二硫键外，蛋白质要行使其功能还需要其他的修饰，包括**切割**和各种**共价修饰**。切割是常见的修饰方式，尤其是利用**氨肽酶**和**羧肽酶**的修剪；但也会发生肽链内部切割，如胰岛素中内部肽段的切除。信号序列通常从分泌蛋白中切除，若当蛋白是多蛋白的一部分时，必须通过切割释放其组成蛋白。**遍在蛋白**是含有头尾相连的多个拷贝的多蛋白，必须被切割以形成单个**遍在蛋白**分子。

化学修饰有多种方式，可以发生在 N 末端、C 末端以及 20 种氨基酸中除 Ala、Gly、Ile、Leu、Met 和 Val 外的大部分氨基酸的侧链上。这些修饰包括**乙酰化、羟基化、磷酸化、甲基化、糖基化**甚至**核苷的添加**。胶原蛋白中 Pro 的羟基化是常见的修饰，而组蛋白常常是被乙酰化（参见 D3）。许多酶的活性，如糖原磷酸化酶及某些转录因子是通过磷酸化来调控的。

蛋白质降解

不同的蛋白有不同的半衰期，调节蛋白需要迅速更新，而且细胞要能够除去有缺陷的和受损的蛋白。在真核生物中，研究发现 **N 端残基**在蛋白稳定性上起关键作用。8 个 N 端氨基酸（Ala、Cys、Gly、Met、Pro、Ser、Thr、Val）显示高稳定性（$t_{1/2}$>20 h），而另 8 个（Arg、His、Ile、

Leu、Lys、Phe、Trp、Tyr）显示短的半衰期（2～30min），而 4 个氨基酸（Asn、Asp、Gln、Glu）经化学修饰后稳定性降低。损伤的、修饰的或者含有不稳定的 N 端残基的蛋白质可通过共价结合高度保守的小分子蛋白质遍在蛋白而被泛酰化，这是通过遍在蛋白分子的羧基端 Gly 连接到蛋白质的 Lys 残基而完成的。遍在蛋白被 26S 蛋白酶复合体（蛋白酶体）消化降解，该反应需要 ATP 并释放完整的遍在蛋白分子供再被利用。

（林会兰　译　刘进元　校）

R1 病毒简介

要 点

病 毒	病毒是极小（20～30nm）的寄生物，无法在宿主细胞外复制、转录或翻译。以细菌为宿主的病毒称噬菌体。病毒颗粒至少由两部分组成：核酸与蛋白衣壳。有些病毒有一个脂蛋白外膜，还有具有为侵染后转录或复制所必需的非结构蛋白。
病毒基因组	病毒基因组由 DNA 或 RNA 组成，可以是双链或单链，而且对于单链基因组来说，可以是正义、负义或双义（根据其 mRNA 序列定义）。基因组的大小可由 1kb 到接近 300kb，复制时利用病毒自身和宿主细胞两方面的酶。
复制策略	病毒的复制策略主要取决于其类型和基因组的大小。小的 DNA 病毒通常能更多地利用宿主细胞的复制系统，而大的 DNA 病毒常编码自己的聚合酶。RNA 病毒的复制则必需有自己编码的依赖 RNA 的聚合酶。有些 RNA 病毒利用依赖 RNA 的 DNA 聚合酶（反转录酶），通过 DNA 中介来复制。
病毒毒性	许多病毒并不引起疾病，病毒毒性的产生机制往往与其生命周期吻合，但也有一些会有助于传播。
相关主题	噬菌体（R2）　RNA 病毒（R4） DNA 病毒（R3）

病毒

很难给病毒下一个准确的定义。病毒这个词最早来源于 18 世纪 90 年代，Jenner 在描述牛痘和天花时把它们认为是一种毒素。**病毒颗粒**（virion）属于亚显微层次，并且只能在宿主细胞内复制。所有病毒的翻译完全依赖于宿主细胞，有些病毒的转录、复制因子也不同程度地依赖宿主细胞。原核生物的病毒被称为**噬菌体**（bacteriophage 或 phage）。

病毒至少由病毒编码的**蛋白衣壳**（capsid）包裹的单链或双链 DNA 或 RNA 基因组组成。衣壳及其与基因组的作用基本决定了病毒的结构(参见图 R1.1，显示不同结构病毒的例子)。

在成熟病毒颗粒形成过程中，病毒衣壳是通过蛋白亚单位逐步组装起来的，该过程可能需要蛋白与基因组的相互作用（基因组与衣壳的复合体称**核壳**，nucleocapsid)。最简单的模型是一些噬菌体（如噬菌体 M13）和较小的哺乳动物病毒（如骨髓灰质炎病毒)，而且一些更大更复杂的病毒也是遵循同样的原则。

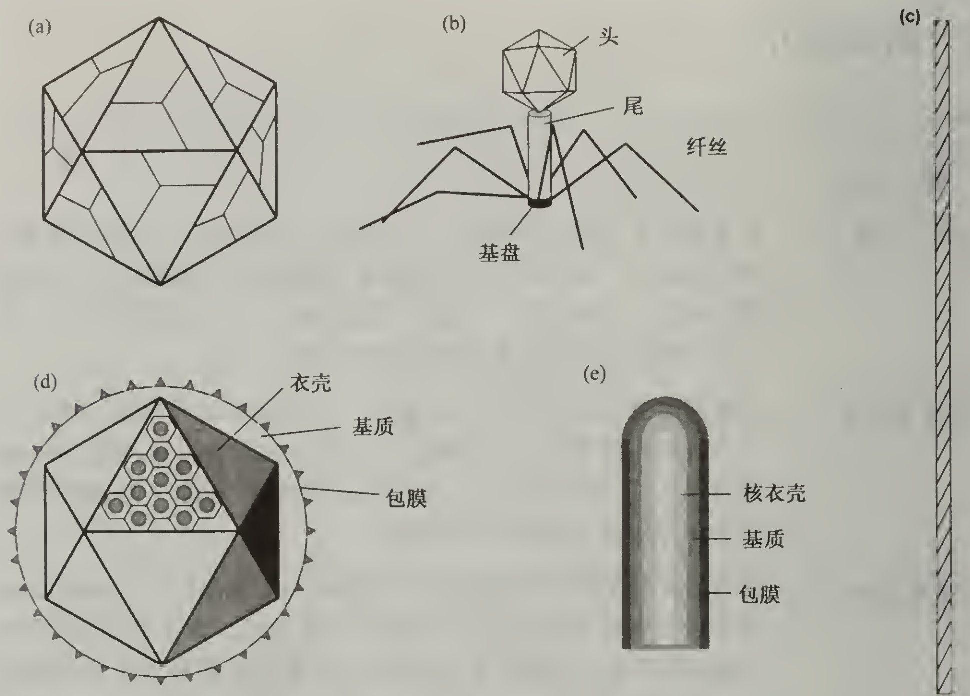


图 R1.1　病毒形态举例（未按比例画）。(a) 二十面体病毒（如骨髓灰质炎病毒）；(b) 有二十面体头部和尾的复杂噬菌体（如噬菌体 T4）；(c) 螺旋病毒（如噬菌体 M13）；(d) 有外膜的二十面体病毒（如疱疹病毒）；(e) 棒状病毒典型的子弹状、螺旋的被膜病毒

许多种病毒有一双层脂蛋白**外膜**。这个外膜来自于宿主细胞膜（通过**出芽方式**，budding），有时会带些宿主细胞蛋白。病毒编码的**包膜糖蛋白**(envelope glycoprotein) 对于病毒的组装与结构很重要，通常是与宿主细胞特殊受体（receptor）结合的**受体配体**（receptor ligand）或**反受体**（antireceptor)。**基质**（matrix）蛋白提供核壳与外膜之间的联系。基质与衣壳蛋白在病毒的转录与复制中可能起非结构作用。结构蛋白往往是重要的抗原，因此可以用来设计疫苗。许多 RNA 病毒和一些 DNA 病毒在病毒颗粒内部也含有一些为感染后即发转录或复制所必需的非结构蛋白。

病毒基因组

病毒基因组是根据在成熟病毒粒内的状态来定义的，并在病毒家族间变异。与真生物（true organism）的基因组不同，病毒基因组可由单链或双链的 DNA 或 RNA 组成。在某些病毒中，基因组仅由一条单一**线状的**(linear) 或**环状的**（circular）核酸分子构成，而在另一些病毒中却可以是**节段型的**（segmented）或二倍体。单链病毒基因组可分成**正义**（positive sense）（核酸序列和 mRNA 相同）、**负义**（negative sense）或**双义**（ambisense，基因可以是正、负义而且经常是重叠编码，参见 P1）。

并非所有病毒都有一个完整的或全功能的基因组，事实上复制出的病

毒粒子数（可通过电镜数）是可侵染病毒粒子数（由细胞培养决定）的100倍以上，甚至常常是数千倍。自身无法复制的基因组可以通过与**辅助病毒**（helper virus）共侵染，借助其具复制能力的**野生型**（wild-type）基因组的生物，或者通过共侵染中基因组的不同突变、或缺失间的**互补**（complementation）（参见 H2）来**恢复**复制能力。

复制策略

病毒的复制或转录策略主要依赖于基因组的类型，不同种群之间差异很大。DNA 病毒通常比 RNA 病毒更多地利用宿主细胞的核酸聚合酶。具有较大基因组的 DNA 病毒如**疱疹病毒**（herpesvirus，参见 R3）比小基因组的如**乳多空病毒**（papouavirus，参见 R3）中的一种猿猴病毒 40（SV40）要更独立于宿主细胞的复制转录机制。RNA 病毒复制所需的依赖于 RNA 的聚合酶一般的宿主细胞中并不存在，需由病毒自己编码。某些 RNA 病毒如**反转录病毒**（retrovirus）编码一种**反转录酶**（reverse transcriptase，一种依赖于 RNA 的 DNA 聚合酶）经 DNA 过渡来复制其自身 RNA 基因组（参见 R4）。

病毒对宿主细胞复制功能的依赖性及对特殊细胞表面受体的要求决定着病毒的宿主细胞特异性。能够提供病毒复制的代谢需求的细胞可以说成是**被允许**（permissive）侵染；而不能提供病毒复制所必需条件的就是**非允许**（nonpermissive）细胞。然而在某些情况下，非允许细胞也可能被侵染，病毒可以对宿主细胞产生显著影响如细胞**转型**（transformation，参见 R3 与 S）。

病毒毒性

有些病毒损坏它们赖以复制的细胞，如果足量的细胞被破坏结果将导致疾病。但病毒并不是为了引起疾病而存在，仅仅是因为它们能够复制。在许多情况下，**毒性**（virulence，引起疾病的能力）是不利的（即毒性会降低病毒自身的复制能力）；而对于另一些病毒来说毒性可能会有助于其传播。毒性的进化往往造成对宿主细胞的破坏能力增强和传播能力的减弱或相反。病毒的毒性机制可分成六大类：

(1) 对细胞代谢的机遇性损伤（如与宿主细胞竞争为复制所必需的酶、核苷酸，或病毒生长因子）。
(2) 在细胞间传播过程中对细胞膜的损伤（例如许多噬菌体使细胞裂解，或疱疹病毒使细胞融合）。
(3) 病症对宿主间传播的重要性（例如普通感冒病毒引起的打喷嚏、狂犬病毒引起的行为改变）。
(4) 逃避宿主的免疫系统，如通过快速突变。
(5) 病毒抗原机遇性的引起有害免疫应答（例如 B 型肝炎病毒）或导致自身免疫病的交叉反应。
(6) 细胞转型（参见 S）和癌变（例如某些乳多空病毒：SV40，参见 R3）。

R2 噬菌体

要 点

一般特性	尽管噬菌体基因组有大有小，生活周期有简单的也有复杂到包括调控病毒与宿主细胞代谢的，但它们都有一个共同的特点即侵染细菌。
裂解性与溶原性侵染	在裂解性侵染中，病毒粒子从溶解细胞中释放出来，而溶原性侵染中病毒则将它们的基因组整合到宿主细胞基因组中，并可能稳定遗传直至转变进入裂解性侵染。
噬菌体 M13	噬菌体 M13 有一个很小的单链 DNA 基因组，以双链 DNA 复制型复制，可进行不引起细胞裂解的侵染。经修饰的 M13 噬菌体已被广泛地用作克隆载体。
λ噬菌体	λ噬菌体可能是研究得最彻底的溶原性噬菌体。通过对不同基因表达的时序调控，可以使该病毒既能进行快速裂解性侵染，又能在环境条件不利的情况下进行作为原噬菌体整合进宿主细胞基因组中的溶原性侵染。*cⅠ* 基因产物即λ阻抑物的表达是确定溶原性侵染的关键步骤。
转座噬菌体	有些噬菌体如噬菌体 Mu，通常整合进宿主基因组，并经重复转座来复制。

相关主题

噬菌体载体（H2） 病毒简介（R1）
转录的起始、延伸与终止（K4） DNA 病毒（R3）
转录调控举例（N2） RNA 病毒（R4）

一般特性

噬菌体（bacteriophage 或 phage）是侵染细菌的病毒。它们的基因组可以是 RNA 或 DNA，大小可以从 2.5kb 到 150kb。它们的生活周期可以很简单或更复杂且高度受控、包括基因组**整合**（integration）甚至**转座**（transposition，参见 F4）。

噬菌体在病毒学和分子生物学历史上扮演着重要角色。Twort 和 d'Herelle分别于 1915 年和 1917 年最初发现噬菌体。噬菌体作为模式病毒而被广泛研究，并在早期的 DNA 作为遗传物质的鉴定、遗传密码的确定、mRNA 的发现以及许多分子生物学的基本概念的建立中成为重要的研究工具。由于噬菌体寄生在原核生物细胞内，它们常含有与宿主极相似的基因组序列，因此作为一个简单模型广为应用于原核分子生物学中的各个方面。有些噬菌体还被用作克隆载体（参见 H2）。细菌学家也常利用种

特异裂解型噬菌体的生物型来研究各种病原细菌的流行。

裂解性与溶原性侵染

有些噬菌体能以惊人的速度复制，侵染、复制、组装以及裂解释放整个过程仅需 20min。在这种情况下，噬菌体基因组的复制独立于细菌基因组。然而有时病毒复制与释放可以不造成溶菌（如噬菌体 M13），而另一些噬菌体则在**裂解性**侵染与溶原性侵染之间转换。裂解性侵染时，DNA 在细胞内复制，**溶原性**侵染则将其 DNA 整合入宿主基因组（λ 噬菌体）。还有一些病毒将基因组整合宿主基因组后以**转座**方式来复制（如噬菌体 Mu）。

噬菌体 M13

噬菌体 M13 有一较小的（6.4kb）环状单链，正义 DNA 基因组（图 R2.1）。其病毒粒子通过位于病毒一端的一种微小衣壳蛋白（g3p）特异性地附着大肠杆菌的性纤毛（由一个称为 F 因子的质粒编码）（参见 A1），这一微小衣壳蛋白的结合诱导了主要衣壳蛋白的结构改变，引起整个病毒粒子变短，并将病毒 DNA 注入宿主细胞。接着宿主酶将其单链基因组转化成单链 DNA **复制型**（replicative form，RF）。M13 基因组有 10 个紧密排列的基因及一个含有复制起点的基因间序列。转录仍旧是利用宿主细胞的酶，可以从几个启动子中的任一个开始，终止于两个终止子中的一个（图 R2.1）。这种转录方式造成靠近终止子的基因比远离终止子的有更大的转录概率，形成了病毒最主要的表达调控方式。RF 的复制是正常的双链 DNA 复制，产生多个 RF 拷贝，但 RF 复制的起始需要病毒内切核酸酶在（+）链上制造一个切割以添加一个 3′-OH，而不是 RNA 引发。最终装配新的噬菌体颗粒经 RF 的连续复制产生众多单股（+）链，与此同时噬菌体基因与蛋白将包裹新生（+）链以阻止互补（-）链的合成。然后这些被包裹的前体被运送至细胞膜上，在那里 DNA 与主要衣壳蛋白结合。同时，新的病毒粒子在不造成细胞裂解的情况下排出细胞膜。被 M13 侵染的细胞仍能继续生长分裂（尽管以低速率），产生出的子代也同样是被侵染的，且能继续释放 M13。另外在各个病毒颗粒内的 DNA 量也变化无常（造成病毒粒子的长短不一），含有多个拷贝基因组的颗粒与仅含有部分基因组的颗粒在任一群体中都会同时并存。上述 M13 生活周期中的几个特征使其成为理想的克隆载体（参见 H2）。双链环形的 RF 在实验中正好可当作质粒操作；而且基因组与病毒颗粒的大小没有严格限制意味着基因组内可以插入较大的外源 DNA 片段；病毒的单链 DNA 基因组使得其 DNA 成为测序反应的理想模板；最后 M13 生活周期的非裂解特性使得获取大量的纯病毒 DNA 的分离变得很容易。

λ 噬菌体

另一个被作为基因表达的调控模型研究得较彻底的噬菌体是 λ 噬菌体，其派生结构常被用作克隆载体（参见 H2）。λ 噬菌体颗粒有一个包裹着 48.5kb 的线性双链 DNA 基因组的二十面体的头部，以及一个长的柔

韧的尾部。该噬菌体能识别大肠杆菌外膜上的特异受体，并通过其尾部将其基因组注入宿主细胞。尽管在病毒颗粒内的基因组是线性的，但它的两端是单链且互补被称作 ***cos* 末端**（*cos* end）（参见 H2）。一旦进入细胞，两个 *cos* 末端就会迅速相互结合，形成有缺口的环状基因组，细胞的 DNA 连接酶可修补该缺口。在感染细胞内，λ 噬菌体既可营**裂解性**的生活周期，也可以是**溶原性**的。在溶原性生活周期内，基因组以线性拷贝或原噬菌体形式整合进宿主基因组。

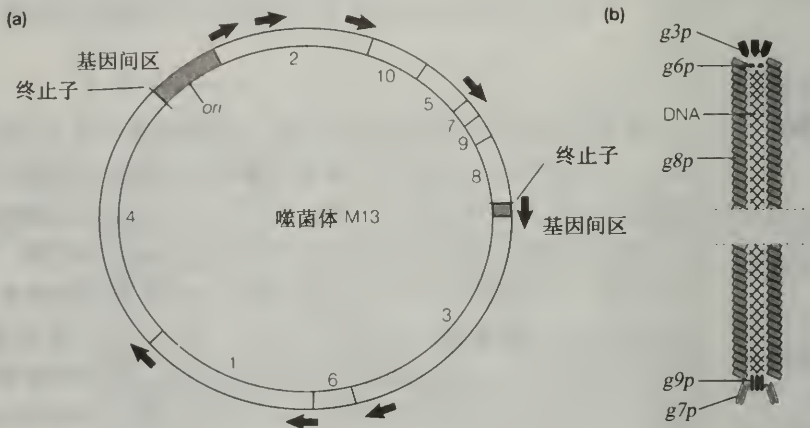


图 R2.1 M13 的基因组（a）及其病毒颗粒结构（b）的示意图。图（a）中的箭头表示启动子，而（b）中的 *g3p* 表示 3 蛋白基因，等等

有三组 λ 基因在侵染后不同时期表达。首先表达的是**立即早期基因**（immediate-early），接着是**晚早期基因**（delayed-early）。这两组基因的表达导致基因组复制。然后**晚期基因**（late）表达产生为组装新病毒颗粒的结构蛋白以及细胞裂解所需蛋白。λ 噬菌体的基因调控机制比较复杂，在此只作简要介绍。图 R2.2 展示 λ 噬菌体基因组。

基因组环化后随即立即早期基因开始转录。转录从 *p*L 和 *p*R 两个启动子起始，并进行两个立即早期基因 *N* 和 *cro* 的转录。这两个启动子利用 DNA 的不同链作为合成 RNA 的模板，分别向左（*p*L）向右（*p*R）转录（图 R2.2）。*N* 和 *cro* 基因转录的终止都依赖 rho 因子（参见 K4）。N 蛋白是一个抗终止子，能使 RNA 聚合酶通读 *N* 和 *cro* 基因的转录终止信号，结果使得转录向左右延伸，右边的转录通过有复制基因 *O*、*P*、进入 *Q*，而左边则转录进相关于重组及复制的基因。这些基因的转录使得宿主酶系统进行 DNA 双向复制成为可能（参见 E4），而在蛋白 pO，pP 与宿主解旋酶（DnaB）的复合体作用下，双向复制从 *ori*（复制起点）位点起始。此后 *gam* 基因产物使复制转化为**滚环复制**（rolling circle replication），产生出线性基因组串联体（多个基因组的串联）（图 R2.3）。

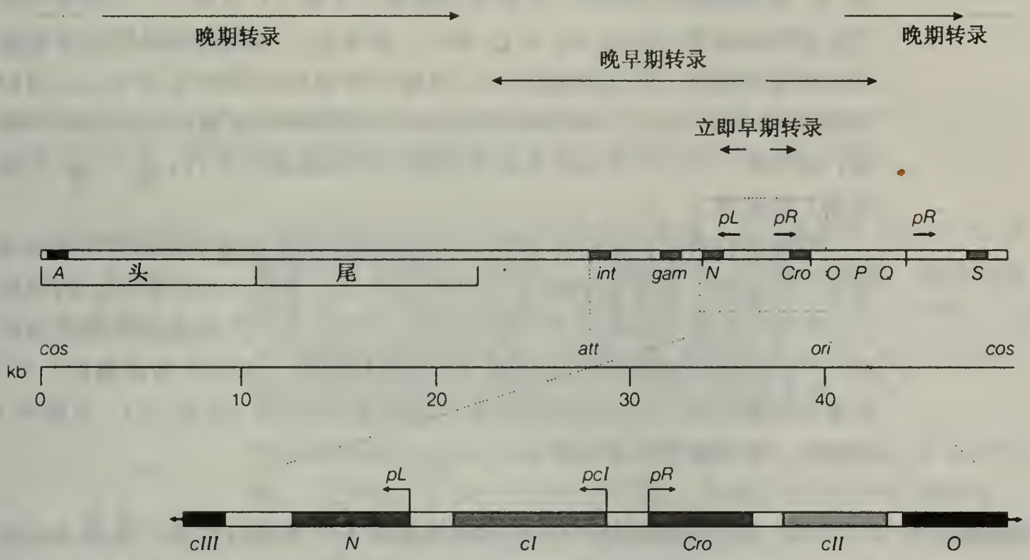


图 R2.2 λ噬菌体基因组的简图（线性化）

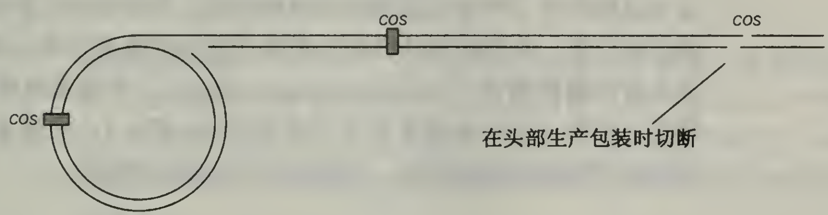


图 R2.3 λ噬菌体的滚环复制

从 *p*R 的转录结果产生出 Cro 蛋白，而该蛋白能结合并覆盖 *p*L 和 *p*R 启动子，抑制从该两个启动子的转录。结果早期转录就被关闭。Q 蛋白与 N 蛋白相似，也具有抗终止功能（可与 HIV 的 Tat 蛋白相比较，参见 N2）。Q 蛋白的生成引起从 *p*R′起始的晚期基因转录的开始（图 R2.2）。晚期基因编码病毒颗粒头部和尾部的结构蛋白、能切割 *cos* 末端产生出线性基因组的蛋白以及导致宿主细胞裂解、病毒释放的蛋白。

整个裂解过程可在 35min 内完成，并能释放出 100 个左右的病毒粒子。裂解循环需要晚期基因的表达，而溶原过程则需要被称为 **λ阻抑物** (lambda repressor) 的合成。阻抑物是 *c*Ⅰ 基因的产物。晚早期基因包括复制和重组基因，还编码三个调控因子。这三个调控因子一个是 Q 蛋白，负责晚期基因的表达，另两个 *c*Ⅱ 和 *c*Ⅲ 调控基因分别由 *p*R 和 *p*L 启动。*c*Ⅲ 产物具有稳定 *c*Ⅱ 产物的作用。*c*Ⅱ 产物能结合并激活 *int* 基因和*c*Ⅰ 基因启动子。上述 *int* 基因负责将 λ 基因组整合进宿主细胞基因组；

而*c*Ⅰ 基因编码λ阻抑物。λ阻抑物抑制 *p*R 和 *p*L 启动子，结果是抑制了所有早期基因（包括 *cro* 和 *Q* 基因）的表达，从而也抑制了晚期基因表达和裂解循环，进入溶原状态。裂解与溶原间的调和取决于 Cro 蛋白（抑制早期基因和 *c*Ⅰ 基因的表达）和利于裂解的 Q 蛋白（激活晚期基因）的浓度，另一个方面也取决于所建立的溶原途径中的 *c*Ⅱ、*c*Ⅲ 的蛋白及λ阻抑物。

溶原性侵染可以维持许多代，在此期间原噬菌体象细菌基因组的其他部分一样复制。溶原期的转录主要受制于 *c*Ⅰ 基因，*c*Ⅰ 有自己的启动子，可由 cⅡ蛋白增强但并不是必需的，因为 *c*Ⅰ 产物也能调控自身转录。一般当感染细胞受到威胁或发生 DNA 损伤（如受到电离辐射）时，溶原平衡被打破。这时将会诱导宿主细胞表达 RecA（参见 F4），以降解λ阻抑物，使病毒进入裂解循环。

转座噬菌体

常在革兰氏阴性菌中发现转座噬菌体，尤其是假单胞杆菌属中的细菌。其中研究最彻底的是噬菌体 Mu（μ）。转座噬菌体与λ噬菌体相似，都有裂解期和溶原期，但它们基因组的复制方式不同。在裂解“早期”，一个由病毒转座酶、细菌 DNA 聚合酶和其他病毒与细菌的酶共同参与的复杂过程介导了病毒基因组的复制和转座，可使其拷贝分布于宿主基因组的任何位置，且原病毒基因组无需离开宿主细胞基因组（参见 F4）。仅经过几轮的**复制转座**（replicative transposition），病毒基因组就被从细胞基因组上切除（且还会带有邻近的细胞 DNA 部分），然后被包装。结果引起宿主细胞基因组的降解，并释放出新的噬菌体颗粒。

R3 DNA 病毒

要 点

DNA 病毒基因组：复制与转录	DNA 病毒的基因组可以是双链或单链。几乎所有的真核 DNA 病毒都在宿主细胞核内复制，且能利用宿主细胞的复制、转录和翻译系统。较大的双链 DNA 病毒往往有比较复杂的生活周期，包括对病毒和细胞的复制、转录、翻译的时序控制，而较小 DNA 基因组的病毒则会更依赖于宿主细胞来完成复制。
较小 DNA 病毒	乳多空病毒科是小 DNA 病毒家族中的一例，如 SV40 和多瘤病毒。它们均只有 5kb 的双链基因组，靠基因重叠和拼接编码 6 个基因。这些病毒能反式激活细胞复制系统，导致病毒和细胞都进行复制，因此会在宿主体内诱发肿瘤。
较大 DNA 病毒	较大的 DNA 病毒中包括疱疹病毒家族。疱疹病毒侵染许多脊椎动物，并能引起多种严重疾病。
单纯疱疹病毒-1	单纯疱疹病毒-1（HSV-1）有 70 个以上的可读框，其基因组长 150kb 左右。侵染细胞后，三组基因即立即早期（α）、早期（β）和晚期（γ）基因会按严格时序表达。这些基因能表达一系列反式激活因子来调控的病毒转录和活化。这种病毒能进行潜伏性感染。

相关主题

真核生物载体（H4） 噬菌体（R2）
真核生物的转录（M） RNA 病毒（R4）
病毒简介（R1） 肿瘤病毒与癌基因（S）

DNA 病毒基因组：复制与转录

较大的 DNA 病毒（如疱疹病毒、腺病毒）具有多至 200 个以上编码基因的双链基因组，其复杂的生活周期不仅包括对自身复制的调控，而且有时还能调控其宿主细胞的生活周期及功能。而在另一极端，乳多空病毒是仅拥有几个编码基因的极小的双链环状基因组，需要依靠宿主来完成大部分复制功能。

大多数真核 DNA 病毒在宿主细胞核内复制，在那里即使最大最复杂的病毒也会利用细胞的 DNA 代谢途径。这意味着病毒和宿主细胞的启动子序列常有相当大的相似性。由于这一原因，DNA 病毒的启动子常被用在哺乳动物表达载体中（参见 H4）。

较小 DNA 病毒

乳多空病毒包括**猿猴病毒 40**（simian virus 40，SV40）。由于它是一种致癌病毒而被广泛研究（参见 S）。SV40 拥有一个 5kb 双链环状超螺旋基因组，并利用来自细胞的组蛋白包装成一个 45nm 的二十面体病毒颗粒。为了在这么小的基因组中编码 5 个基因，结果是两条链均编码并且相互还有基因重叠。这些基因可以分成两个重叠的转录单元：早期和晚期基因（图 R3.1）。通过不同的重叠可读框及不同的拼接方式可产生。SV40 依靠宿主细胞的酶来转录、复制，其早期基因可产生转录活化因子（称为**大 T 抗原**和**小 t 抗原**，large T-antigen and small t-antigen）来激活病毒和宿主细胞的转录与复制。这就是该病毒致癌的原因。晚期基因能产生三个为病毒颗粒包装所需的衣壳蛋白 VP1、VP2 和 VP3。

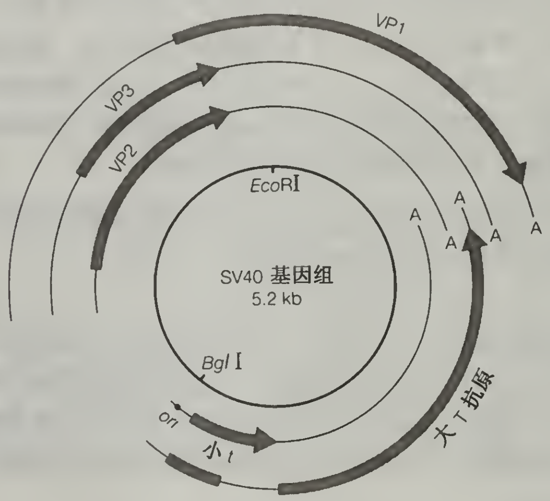


图 R3.1 SV40 病毒基因组的结构。外侧线表示转录物和编码区，A = poly（A）尾部

较大 DNA 病毒

疱疹病毒是“大”DNA 病毒与宿主之间复杂的相互作用的最好例子。疱疹病毒感染脊椎动物，其病毒粒子很大、具二十面体及被膜（参见 R1），拥有一条 270kb 以上编码 100 个左右可读框（ORF）的双链线性 DNA 基因组。根据其生物学特征及基因组组构可将疱疹病毒分成三个亚组，总共约有 100 多种，而且大部分是宿主专一性的。其中感染人类的病毒包括：能引起冻疮的单纯疱疹病毒、能引起水痘的水痘带状疱疹病毒、能引起传染性的单核细胞增多症（腺热）及某些肿瘤的非洲淋巴瘤病毒。

单纯疱疹病毒-1

HSV-1 已被研究得很透彻。它的基因组大小约为 150kb，含有 70 多个 ORF，DNA 的两条链都编码基因，有时还会相互重叠，并且可以将整个基因组（图 R3.2）分成长短两个部分：每一部分都含有一个单一片段（U_L 和 U_S），两个片段之间是反向重复序列（IR_L 和 IR_S），在两端也有未

端反向重复序列（TR_L 和 TR_S）。这些反向重复由 b、b′和 c、c′组成。另外还有一个短序列（a）在其中重复出现，重复次数不定（a_n 和 a_m）。

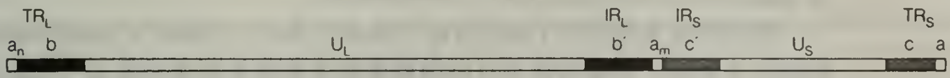


图 R3.2 HSV－1 的基因组结构

疱疹病毒基因组的转录与复制受严格的时序控制。在感染**允许细胞**（permissive cell，允许病毒侵染和复制的细胞，参见 R1）后，基因组环化位于末端重复区域间的一组基因，**立即早期基因**或称为 **α 基因**在 RNA 聚合酶Ⅱ的作用下开始转录，该 α 基因的转录是被病毒编码蛋白即 **α 反式诱导因子**（α-*trans*-inducing factor 或 α-TIF）反式激活的。与 HSV－1 基因产物中约三分之一一样，α-TIF 并不为复制所必需，而是作为成熟病毒颗粒基质的一部分，在进入细胞后与细胞转录因子发生作用，并与 α 启动子上游的特定序列结合，从而提高它们的表达。α mRNA（其中某些 mRNA 经受拼接）编码**早期基因**或 **β 基因**的反式激活因子，而 β 基因又编码大多数用于进一步转录与基因组复制所需的非结构蛋白及少数结构蛋白。早期基因产物包括参与核苷酸合成所需的酶（如腺苷激酶、核苷酸还原酶）、DNA 聚合酶、立刻早期基因表达的抑制蛋白以及能够降低宿主细胞各方面代谢的调节蛋白。β 基因的启动子与它们宿主的启动子很相似，蛋白均在转录启动子点上游 20～25bp 处有明显的 TATA 框及 CCAAT 框，以及转录因子结合位点。疱疹病毒基因组整合到宿主基因组中后，β 基因启动子能很好地启动转录，因而早就被当作模式 RNA 聚合酶Ⅱ启动子被研究。

复制可从环状基因组上几个可能的起始（ORI）位点中的任何一个开始，需要有病毒所编码的 ORI 结合蛋白、解旋酶-引发酶复合体和聚合酶-DNA 结合蛋白复合体的参与。DNA 合成是半不连续的（参见 E），产生多个复制复合体和复制叉构成的多联体。

晚期基因或 **γ 基因**（有些只能在 DNA 复制后才表达）主要编码结构蛋白，或者是侵染后立即需用的因子（如 α－TIF）。该病毒在核内组装，空的衣壳与游离基因组末端的重复“a”序列结合，接着装入一个基因大小的序列，然后切割再从下一个“a”序列处开始重复包装循环。病毒的外膜来自修饰的核膜，并含有为附着与进入所需的重要的病毒糖蛋白。

除了上面所说的严格受控的复制循环外，疱疹病毒还能进行**潜伏性感染**（latent infection），它们能下调自身的转录至可使环状基因组在宿主细胞核内保持外源染色体状态而不进行复制的水平。潜伏性感染可在整个细胞生存期持续，在此期间可能会有周期性的复制。HSV－1 主要在神经元内潜伏，而其他疱疹病毒潜伏侵染包括淋巴细胞在内的其他组织。疱疹病

毒潜伏与复制的精确调控机制还没有完全搞清楚，可能与编码在末端重复区内和某些 α 基因重叠并互补的一些特殊 RNA（潜伏关联转录物，latency-associated transcripts，LAT）的转录相关。

疱疹病毒及其他大 DNA 病毒（如腺病毒），凭借巨大的基因组和复杂的生活周期，常编码许多只对理想复制重要并非必需的基因，尤其是在细胞培养期（如腺苷激酶基因）。与较小的病毒相比它们更能克服基因组大小的巨大变异，而不失去稳定性。这些特性（以及重组 DNA 比 RNA 相对容易）使得它们用作实验室进行外源基因操作的载体以及用作疫苗。

R4 RNA 病毒

要 点

RNA 病毒基因组的一般特性

病毒 RNA 基因组可以是单链或双链、正义或负义，而且还有多种不同的复制机制。由于完全是依赖病毒编码的依赖于 RNA 的聚合酶来进行复制，其合成互补 RNA 链的错误率要远远大于依赖于 DNA 的聚合酶。这在很大程度上影响着 RNA 病毒的进化，这种进化朝着增强承受能力的方向发展，但却限制了 RNA 病毒基因组增大。

病毒的反转录

来自病毒的反转录酶（RT）的利用是分子生物学领域划时代的革新。尽管不同的 RNA 病毒家族复制方式大不相同，但均需经反转录来完成其基因组的复制，它们反转录酶的相似足以暗示它们来源于同一祖先。

反转录病毒

反转录病毒有二倍体的、正义链 RNA 基因组，经由双链 DNA 中介完成复制。这个中介称作原病毒，能够插入到宿主细胞基因组内。反转录病毒与真核反转录转座子如酵母 Ty 元件有许多相似的特性。

致癌性反转录病毒

反转录病毒对宿主基因组的插入可引起宿主基因失控，或有时会引起病毒与宿主基因重组（病毒获得宿主的基因）。如果反转录病毒改变了一个重要的细胞调控基因称为癌基因的表达或活性就可能导致癌症。

反转录病毒基因组结构及其表达

反转录病毒基因组的基本结构是 *gag*、*pol* 及 *env* 基因两侧排列 5′及 3′-长末端重复序列（LTR）。负责整个病毒转录的启动子存在于 5′- LTR 的 U3 区域。病毒的转录产物会被加上 poly A 尾，而且可能需要拼接。在人类免疫缺陷病毒（HIV）中，Tat 调控从病毒启动子开始的转录延伸；而 Rev 调节将未拼接的 RNA 向细胞质的运送。

反转录病毒的突变率

某些反转录病毒的反转录酶复制错误率可高达万分之一。有缺陷的基因组可以通过互补与重组来得到挽救，因此结合病毒快速变异（对 HIV 来说，每天有 $10^9 \sim 10^{10}$ 个新病毒产生），能使它们很快适应选择压力。

相关主题

cDND 文库（I2）　　病毒简介（R1）
RNA 聚合酶 Ⅱ 基因：启动子与增强子（M4）　　DNA 病毒（R3）
转录调控举例（N2）　　肿瘤病毒与癌基因（S）

RNA 病毒基因组的一般特性

依据病毒家族，病毒 RNA 基因组可以是单链或双链，若是单链基因组则有正义与负义之分。由于宿主细胞没有依赖于 RNA 的 RNA 聚合酶，所以只能由病毒自己编码，且该聚合酶基因（*pol*）(该基因还常编码其他非结构功能蛋白）通常是 RNA 病毒基因组中最大的基因。这使 RNA 病毒成为真正的翻译寄生者，而复制和转录往往完全独立于宿主细胞核，因此与真核 DNA 病毒不同，许多 RNA 病毒在细胞质中复制。

依赖于 RNA 的聚合酶不像依赖于 DNA 的聚合酶得那样准确，通常 RNA 病毒无法进行校正（参见 F1），因此许多 RNA 病毒有 $10^{-3} \sim 10^{-4}$ 的突变率（即在每个复制循环中每 $10^3 \sim 10^4$ 个碱基中有一个突变），而大 DNA 病毒中却可以低到 $10^{-8} \sim 10^{-11}$（如疱疹病毒）。这里有三个主要结论：

(1) RNA 病毒变异率很高，因此如果该病毒复制周期快，那么病毒的抗原性与毒性就会很快地发生较大的变化（即 RNA 病毒能很快进化以适应改变了的新环境、新宿主等）。

(2) 有些 RNA 病毒变异速率太快以致于它们以**准种**（guasispecies）形式存在于单一宿主内，即不同基因组的群体（常进行互补复制），而且只能根据大部分或平均序列对它们给出分子上的定义。

(3) 由于大多数突变对病毒复制是有害的，因此突变率给 RNA 病毒基因组的大小设定了一个约 10^4 碱基的上限，突变率的倒数（即每个基因组平均发生一个突变的大小）。

病毒的反转录

反转录过程以及**反转录酶**（reverse transcriptase，RT）的利用在上文中已有叙述（参见 I2)。尽管采用反转录的不同病毒会有不同的形态、不同的基因组结构和生活周期，但它们反转录酶及其核心蛋白的序列是高度相似的，足以推测它们进化自同一祖先。反转录病毒基因组与许多真核细胞中所发现的反转录转座子如酵母的 Ty 反转录转座子在结构上有明显的相似（序列同源）(参见 F4)。

反转录病毒

反转录病毒有一条单链 RNA 基因组。每一病毒粒内有两个拷贝的正义链基因组。病毒侵染细胞后，单链 RNA 由反转录酶转化成双链 DNA 拷贝，随后的复制与转录以这一 dsDNA 中介体为模板进行。这一中介体通常称为**原病毒**（provirus)，可通过病毒**整合酶**（integrase enzyme）整合进宿主基因组。反转录病毒复杂程度不一（如图 R4.1)，极简单的基因组仅只比反转录转座子多一个侵染所需的包膜糖蛋白基因（*env*)。而另一极端，慢病毒属像**人类免疫缺陷病毒**（human immunodefeciency virus, HIV）具有编码能在复制期的不同阶段起调控作用的各种顺式作用元件及反式作用因子的巨大基因组，并具有调控病毒和细胞功能的能力。

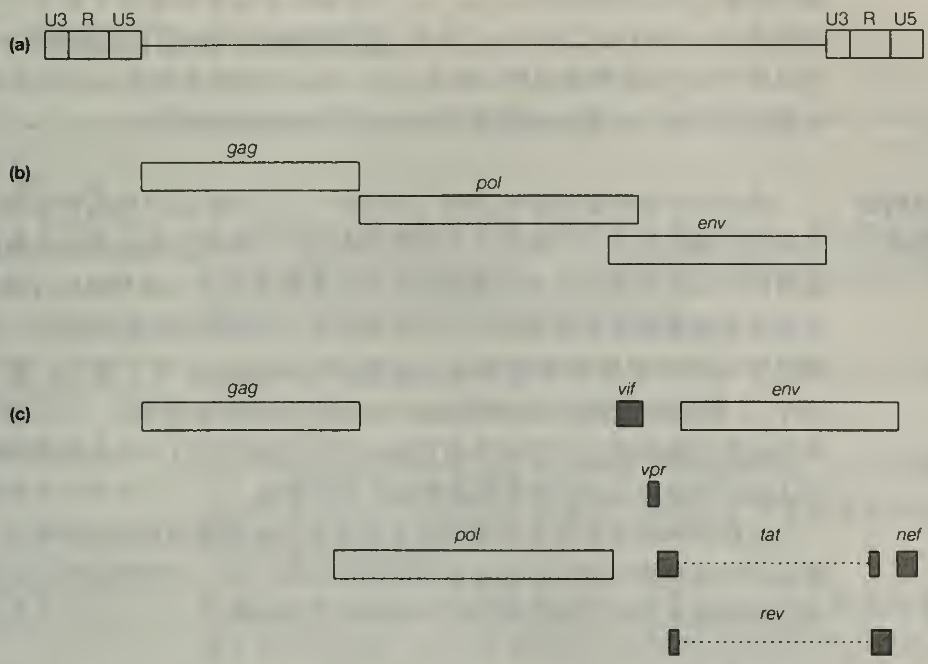


图 R4.1 反转录病毒原病毒的基因组结构。(a) 反转录病毒原病毒的基本结构，含有由 U3、R 和 U5 元件组成的长终止重复 (LTR)；(b) 禽类白血病病毒 (ALV)，一种简单的反转录病毒；(c) HIV-1，一种含有附加调控因子的复杂反转录病毒

致癌性反转录病毒

反转录病毒有时会与宿主基因发生重组，因此这些病毒通常有复制缺陷。反转录病毒会干扰宿主基因的表达调控，或者会将宿主 DNA 片段插入到自己的基因组内实现重组。如果该宿主细胞基因编码一个调控细胞分裂的蛋白，就会引起癌症，这个基因也就是**癌基因** (oncogene) (参见 S)。目前已经鉴定许多表达人癌基因的致癌性反转录病毒。

反转录病毒基因组结构及其表达

反转录病毒基因组结构的几个特例如图 R4.1 所示，可见所有的反转录病毒都有相似基本结构。*gag* 基因编码二十面体衣壳的核心蛋白，*pol* 基因编码病毒复制相关酶（如反转录酶、RNase、RNaseH、整合酶及蛋白水解酶），而 *env* 基因编码包膜糖蛋白。在病毒基因组的两端分别是与复制、宿主细胞整合和病毒基因表达相关的单一元件（U5 及 U3）和重复元件。

整合后的原病毒的转录要依赖宿主 RNA 聚合酶Ⅱ，由 5′端长末端重复序列（LTR）中的 U3 区的启动子启动。RNA 转录产物被加上 poly A 尾，且可能是经拼接的产物。未经拼接的 RNA 在细胞质核糖体上通过诸如读框位移方式（核糖在模板 RNA 上滑过一个读框以避开一个终止密码子）翻译产生 gag 多聚蛋白和 gag-pol 多聚蛋白。拼接后的 mRNA 在与膜结合的核糖体上翻译产生包膜糖蛋白。有些反转录病毒如慢病毒属的病毒

能产生多个或不同拼接的 RNA，它们能翻译出各种调控因子诸如分别调控转录与 mRNA 加工的 Tat（参见 N2）和 Rev。Tat 蛋白调控从 HIV 基因组的 5′ LTR 的转录延伸（参见 N2）。Rev 可促进未拼接 mRNA 从核向细胞质的转运，以便在细胞质中合成充足的病毒结构蛋白。

反转录病毒的突变率

某些反转录病毒的反转录酶（如 HIV－1）约有 10^{-4}的高错误率，即平均每个基因组有一个突变（上文已有叙述）。结果复制出许多有缺陷的基因组。虽然可通过与另一基因组的互补复制来克服上述种缺陷（通常一个反转录病毒颗粒包装有两个 RNA 基因组）。但在反转录期间任一病毒颗粒中的两个不同的基因组会发生重组。综合以上两个特点，再加上 HIV－1 病毒的快速更新（每天会产生 10^9～10^{10}个新病毒），足以使它能很快地适应新环境（如对抗体或药物处理的选择压力），这也是逐渐被人们认识了的诸如 AIDS 病等感染病理的关键特征。

反转录病毒具有将自身基因组整合进宿主细胞基因组的能力以及相对容易对其复制能力进行实验室遗传修饰的特性。这二个特性使得它们成为创造新细胞系及在生物基因治疗中使用的主要载体。

（陆　韵　译　刘进元　校）

S1 肿瘤病毒中的癌基因

要 点

癌 症	癌症是阻断正常细胞生长的调控机制突变的结果。正常细胞的生长经受多种不同类型的调控，而在正常细胞逐步转变成恶性肿瘤细胞的过程中这些调控可相互独立地丧失。
癌 基 因	癌基因是那些活性过高的能使细胞癌化的基因。相对于未突变（正常）的基因，癌基因是处于遗传显性状态。
致癌性反转录病毒	第一个癌基因是从致癌性反转录病毒中分离得到的。反转录病毒通过表达细胞生长调控的突变基因，或激化正常细胞基因的过量表达而导致细胞癌化。
癌基因的分离	癌基因的分离可以借助于检测 DNA 转变 NIH－3T3 鼠成纤维细胞生长方式的能力的测活来进行。这种测法有很大的实用价值，已用来分离出许多癌基因，但其应用也是有局限性的。
相关主题	诱变（F1）　　癌基因的分类（S2） RNA 病毒（R4）　　肿瘤抑制基因（S3）

癌症

癌症是正常细胞生长失控而造成的疾病。正常细胞的生长发育受到多种不同方式的调控。**完全恶性**（malignant）的肿瘤细胞就可能丧失了全部或是大部分的调控。然而只有当丧失了某些调控出现中间态细胞时才可被检测到。从正常细胞到恶性肿瘤细胞的转化是一个**多步过程**（multistep process），每一步对应于一个正常细胞调控机制的瘫痪。

人们早就认识到癌症是与基因相关的疾病。现已有三方面的证据：

- 发展成某一特定类型的癌症是可遗传的；
- 在某些癌症患者中，肿瘤细胞的染色体明显不正常；
- 能引起突变的物质往往也能引起癌症（参见 F1）。

正常生长的失控好像是由于负责调控的基因发生了突变。癌症也就是在肿瘤细胞内一系列特殊的突变的积累造成的结果。近年来，对癌化过程中突变基因的鉴定取得了长足的进展。引起癌症被鉴定的第一个基因被命名为**癌基因**（oncogene）。

癌基因

癌基因就是那些表达后能引起细胞癌化的基因。这种基因的正常状态称为**原癌基因**（proto-oncogene）不引起癌化，一旦发生突变就异常活跃，其遗传显性程度远远高于原癌基因，以致于只要有一个癌基因就足以改变

细胞行为。

致癌性反转录病毒

癌基因的基本概念是从致癌性病毒，尤其是**致癌性反转录病毒**（oncogenic retrovirus）的研究中得来的（参见 R4）。致癌性病毒是动物癌症的主要病因，而对于人类仅有少数种类的癌症与病毒有关。反转录病毒是 RNA 病毒，通过 DNA 中介（**原病毒**）复制，并能插入到宿主基因组中，借助细胞的 RNA 聚合酶转录病毒 RNA。许多致癌性反转录病毒中有一个多余基因，而具有亲缘关系的非致癌性病毒中却没有。通过将该基因转染到末癌化的细胞后能致使该细胞癌化的实验表明该基因就是癌基因。不同的致癌病毒有不同的癌基因。

癌基因的分离

癌基因的分离与其他生化分子的分离一样，依赖于特殊测定法的有效性。分离癌基因的主要方法是应用对 **NIH-3T3** 细胞的 DNA 转染实验来进行。NIH-3T3 是一个永久的鼠成纤维细胞系（在体外生长良好的结缔组织细胞），将其注入免疫缺陷的小鼠不会引发肿瘤。培养中的 NIH-3T3 细胞生长模式是正常的非癌化细胞的模式。将待测 DNA 转染 NIH-3T3 细胞（即将 DNA 细小沉淀加到细胞上），结果少量细胞会吸收并表达外源 DNA。如果进入的 DNA 含有癌基因，则培养细胞的生长模式就会呈现出癌细胞的特征（图 S1.1）。这一测定法的优点在于：

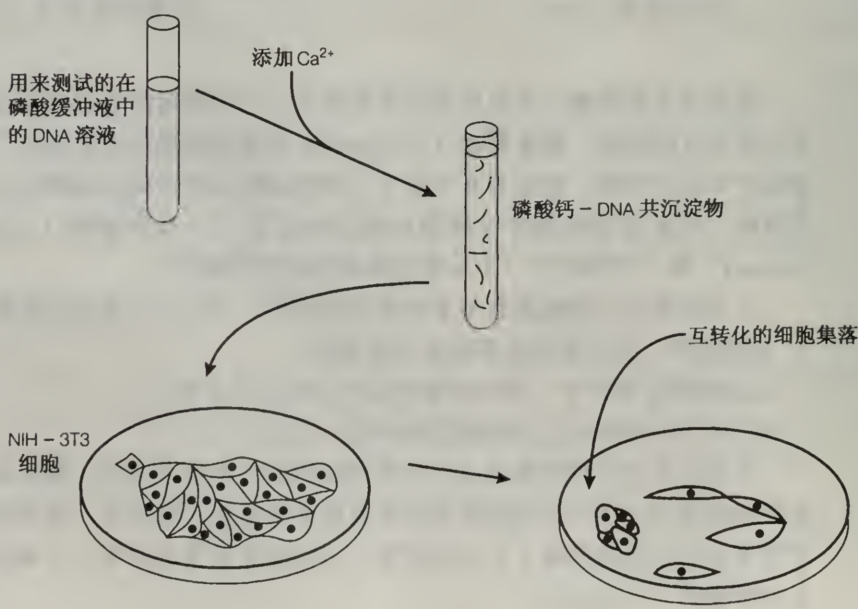


图 S1.1 通过检测引起 NIH-3T3 细胞的生长模式的改变与否来确定被转染的 DNA 中是否存在癌基因

- 测定只需要进行细胞培养，而不是整个动物试验，所以特别适合大量

的样品筛选；

- 与体内试验相比能更加快地获得试验结果；
- NIH－3T3 细胞很易吸收及表达外源 DNA；
- 与体内试验相比这是一个技术比较简单的测法。

与体内试验相比操作更简单，但在广泛的应用中也暴露出许多实际的与潜在的弊端：

- 某些癌基因是细胞专一性的，所以不可用鼠成纤维细胞来测定。
- 因为很难对大基因进行完整转染，大基因会容易丢失。
- NIH－3T3 细胞是一个永久细胞系而不是“正常”细胞，所以致癌早期相关基因可能检测不到。
- 该方法要求被转染的基因呈遗传显性状态，因而将无法检测肿瘤抑制基因（参见 S3）。

第一个癌基因是从致癌性反转录病毒中分离出来的。当反转录病毒的癌基因被克隆，且被用作杂交探针时，取得了一个重要发现，即正常细胞的 DNA 中也有癌基因的同源序列。后来又发现反转录病毒的癌基因起源于正常细胞中的原癌基因，可能是因为当原病毒整合进正常细胞的原癌基因附近时被掺进病毒基因组的。随后，从非病毒感染的癌症患者身上分离出相似（有时是相同）的癌基因。在所有的情况下，癌基因与正常的原癌基因在许多重要方面有区别。

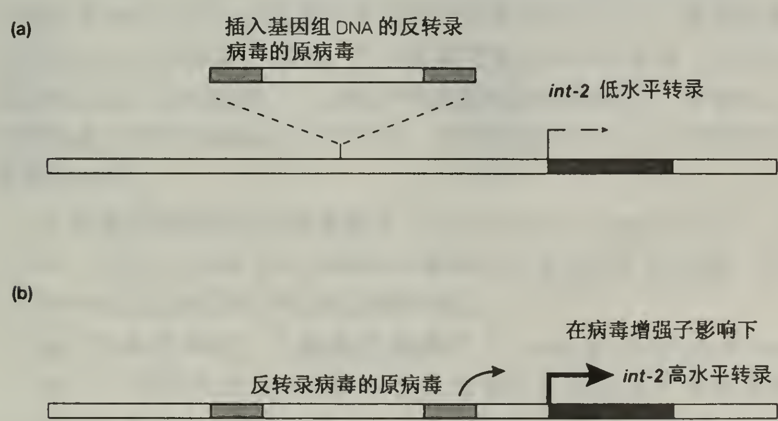


图 S1.2　MMTV 引起鼠乳腺癌的机制。(a) MMTV 原病毒被整合前的鼠 *int*－2 基因；(b) 原病毒的整合导致 *int*－2 的过量表达

- **量的区别**（A quantitive difference）。基因的编码功能可能没有被改变，但由于诸如受控于病毒的启动子或增强子，或是它被转移到基因组的另一新位点，致使基因以高速率或在不正常环境下转录，结果造成正常基因产物的过量形成。例如，鼠腺癌是由**鼠乳腺瘤病毒**（mouse mammary tumor virus，MMTV）引发的，即当 MMTV 原病毒插入到

鼠 *int*－2 基因（编码生长因子）附近，病毒的增强子序列就会强烈刺激 *int*－*2* 的活性（图 S1.2），从而产生出过量的生长因子致使细胞不停地分裂的结果。

- **质的区别**（A qualitative difference）。编码序列也可能会被改变，如经基因缺失或点突变，导致基因产物在功能上的改变，常常是变得过分活跃。例如 ***erbB*** 癌基因编码一个被截短了的生长因子受体，由于缺失的部分正好是生长因子的结合部位，导致该癌基因呈现组成性活性，不停向核（图 S1.3）发送信号，指导细胞不断分裂。

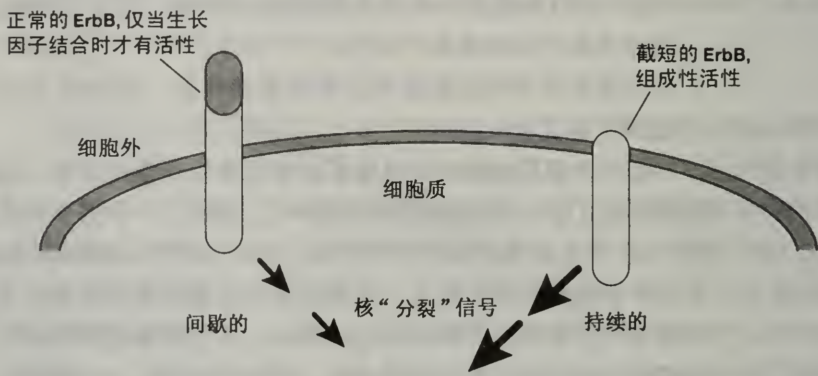


图 S1.3　*erbB* 癌基因编码具有组成活性的生长因子受体

S2 癌基因的分类

要 点

癌基因与生长因子	许多癌基因编码蛋白参与细胞生长因子调控过程中的各个步骤，由这样的癌基因所致变的癌细胞就成了好像在被生长因子不停地刺激而不断进行分裂的细胞。
核癌基因	另一些癌基因编码核 DNA 结合蛋白，来作为转录因子调控与细胞分裂相关的基因表达，其结果导致在癌细胞中细胞分裂相关基因表达的调控机制失效。
癌基因间的协同作用	用癌基因转染直接取自鼠正常成纤维细胞时发现，生长因子相关的癌基因及核癌基因二者均为细胞完全癌化所必须。这反映出体内细胞的癌化过程至少需要一个以上的变化步骤。
相关主题	真核生物的转录因子（N1）　　肿瘤病毒中的癌基因（S1）

癌基因与生长因子

癌基因的分离使得研究每个癌基因在癌化过程中的作用成为可能。许多癌基因编码生长因子应答机制中各个步骤相关的蛋白。**生长因子**是经许多种细胞分泌到体液中的肽，它们能够与附近的组织细胞（可以是与分泌细胞相同或不同的细胞）上的特异性**细胞表面受体**结合，并激起一个通常导致细胞分裂速率增加的应答反应。表现为依赖于其生长因子相关活性的癌基因包括：

- *sis* 癌基因编码**血小板生长因子**（platelet-derived growth factor，PDGF）的一个亚基，如果含有该基因的细胞也含有 PDGF 的受体，则该生长因子的过量表达就会自动促进癌细胞生长。
- *fms* 癌基因编码一个**集落刺激因子 1**（colony-stimulation factor - 1，CSF - 1)受体的突变型，该因子在血细胞形成过程中刺激骨髓细胞。在 Fms 蛋白中，正常 CSF - 1 受体 C 端的 40 个氨基酸被 11 个不相关的氨基酸所代替，结果 Fms 不管 CSF - 1 存在与否，Fms 蛋白总处在活性状态（图 S2.1）。
- 各种 ***ras*** 癌基因编码质膜蛋白中的 **G 蛋白**（G-protein）家族的成员。G 蛋白能将来自许多细胞表面受体的信号传递给酶，从而产生**第二信使**(second messenger)。正常的 G 蛋白被激活时能结合 GTP，而自身的 GTP 酶活性也可使自身失活。*ras* 癌基因的点突变抑制了自身的 GTP 酶活性，结果它们可比正常蛋白维持更长活性（图 S2.2）。

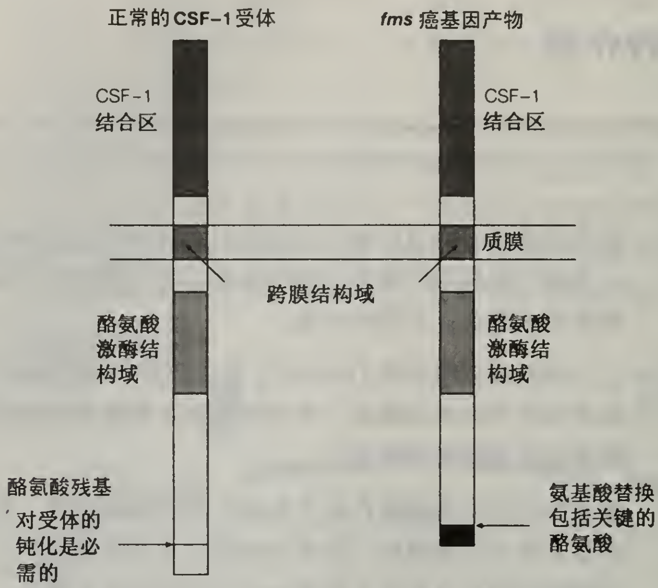


图 S2.1 *fms* 癌基因编码一个突变成组成活性的生长因子受体

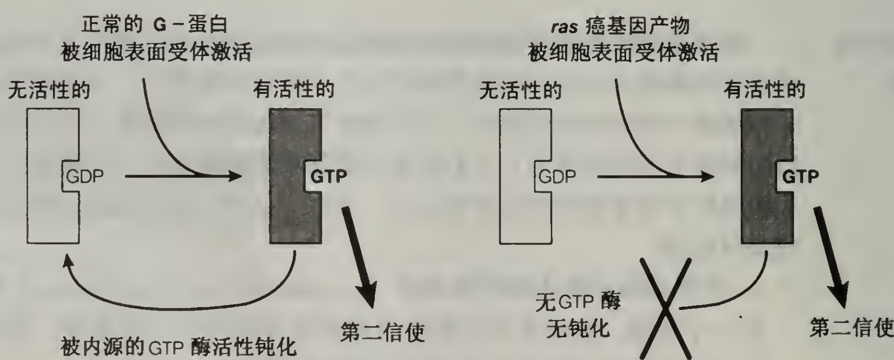


图 S2.2 *ras* 癌基因编码的信号转导蛋白发生突变以致失去了钝化自身的能力

核癌基因 另一类癌基因编码作为转录因子（参见 N1）调控其他基因的表达的核 DNA 结合蛋白。

- 正常细胞中 *myc* 基因的表达是由多种促细胞分裂剂（促进细胞分裂的试剂）诱导的，其中包括 PDGF。*myc* 编码的蛋白能与特殊的 DNA 序列结合，并激活为细胞分裂所需基因的转录。癌细胞中 *myc* 的过量表达可受多种机制作用，可以是病毒增强子导致的转录量增加，也可以是其编码序列从 8 号染色体的正常位置易位到 14 号染色体，受控于免疫球蛋白重链的活性启动子之下所造成的表达增加，或者还可以是 mRNA 的 5′端非编码序列的缺失而引起 mRNA 半衰期延长。
- *fos* 和 *jun* 癌基因编码正常转录因子 **AP-1** 的亚基。在正常细胞中，

fos 和 *jun* 的表达是很短暂的，紧跟在促细胞分裂剂刺激后；而 *fos* 和 *jun* 的基因产物的正常细胞浓度不仅被基因转录效率，而且还被mRNA稳定性来调控。而在癌细胞中，这两过程都被增加了。

- ***erbA*** 是禽类成红细胞增生病毒中发现的第二个癌基因（除了 *erbB* 外）。该基因编码一个甲状腺激素核受体的缺陷型。**甲状腺激素受体**在与甲状腺激素结合后被激活，作为调控特殊基因表达的转录因子。ErbA蛋白缺少正常受体的 C 端序列，因而不能与激素结合，也不能激活基因转录，但它仍旧能与 DNA 的相同位点结合，从而成为正常甲状腺激素受体的拮抗物。

癌基因间的协同作用

正常细胞向完全癌化细胞的转变是一个多步骤的过程，涉及到多个基因表达的改变。尽管在大多数情况下，将上述单个癌基因转染将导致 NIH－3TC 细胞系细胞的癌变，而当用直接从鼠身上取出的成纤维细胞的培养物来代替该细胞系则会出现不同结果。仅有 *ras* 或仅有 *myc* 癌基因都无法使正常细胞完全癌化，只有当两癌基因的同时转染才能达到完全癌化。其他还有多种癌基因的配对能够引起正常鼠成纤维细胞癌化，而单独基因作用同样不行。有趣的是，若要有效的话，配对的两个癌基因必须与一个是生长因子相关，另一个是核癌基因。所以任何一个有活性的癌基因只能完成将正常细胞完全转变成癌化过程所需变化中的一部分。

S3 肿瘤抑制基因

要 点

概述	肿瘤抑制基因由于突变而失活将引起细胞癌化。在正常细胞内肿瘤抑制基因起限制细胞分裂速率的作用。
肿瘤抑制基因存在的证据	肿瘤抑制基因存在的证据是多样的且间接的，其中包括正常细胞与癌细胞融合后的杂交细胞的行为、某些家族性癌症的遗传特征以及作为肿瘤细胞染色体标记的“**杂合性的丢失**”。
RB1 基因	是第一个被分离到的肿瘤抑制基因，已被证明是引起儿童期眼瘤，即成视网膜细胞瘤。*RB1* 基因的突变已在乳腺癌、结肠癌及肺癌患者中被检测到。
P53 基因	*P53* 是多种不同肿瘤中发生突变的肿瘤抑制基因。最初被鉴定时发现该基因既有癌基因的特征，又有肿瘤抑制基因的特征，现在已知它是一个以**显性失活**（dominaut-negative）方式干扰剩下的正常等位基因功能的肿瘤抑制基因。
相关主题	肿瘤病毒中的癌基因（S1）　癌基因的分类（S2）

概述

肿瘤抑制基因与癌基因在作用方式上根本不同。原癌基因通过突变转变成活性大大增加的癌基因；而肿瘤抑制基因转变成致癌性则是发生消除其正常活性的突变的结果。正常的未突变的肿瘤抑制基因能够阻止进入有丝分裂，从而抑制正常细胞分裂，而将这一负调控除去就可恢复细胞分裂。这种作用机制的一个重要后果是为了移去所有限制，肿瘤抑制基因的两个等位基因均失去活性，这正是肿瘤抑制基因以隐性遗传方式起作用的证据。

肿瘤抑制基因存在的证据

不像癌基因有 NIH－3T3 测定法那样，对肿瘤抑制基因目前还没有快速方便的测定法，只有几个肿瘤抑制基因得到分离，且在很长一段时期内这类基因在癌化过程中的根本重要性一直没有得到鉴定。现在已知它们的确存在，而且发现了越来越多的相关证据。

- 早在 19 世纪 60 年代，就已发现正常细胞与癌细胞（不同种的）融合所产生的**杂交细胞**（hybrid cell）一律是非癌性的，经传代，杂交细胞失去染色体后，会转变成癌细胞表型。转变常常与某个特定细胞正常染色体（携有肿瘤抑制基因）的丢失有关。
- 检查某些**家族性癌症**（familial cancer）的遗传特征发现癌症由隐性突

变造成。

- 在许多癌细胞内，总是有某一特定染色体的特征区域的丢失。这种**"杂合性的丢失"**（loss of heterozygosity）表明在丢失染色体片段上的肿瘤抑制基因的丢失图 S3.1。

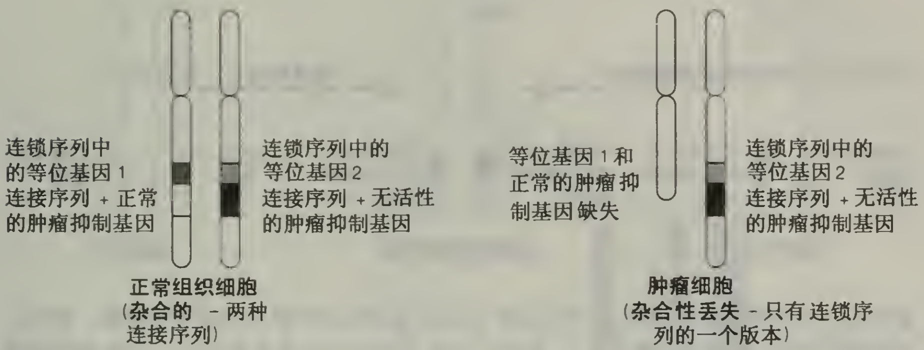


图 S3.1 杂合性的丢失暗示出细胞丢失了含有惟一肿瘤抑制基因活性的等位基因的染色体部分的过程

成视网膜细胞瘤（retinoblastoma）是一种儿童眼瘤，是由于肿瘤抑制基因丢失造成癌症的可靠例证。成视网膜细胞瘤有家族性的（占 40%）和散发性的两种形式。家族性形式是以隐性基因方式遗传，常常是双眼发病；散发型并不经家族遗传，通常只有一只眼睛发病。上述结果暗示出成视网膜细胞瘤是由同一基因的两个等位基因均失活的两个突变造成的。在家族性发病形式中，一个已突变的等位基因是种系遗传的，它自身是无害的，而当剩下的另一个正常等位基因也发生突变时就会在成视网膜细胞中引发肿瘤。在每只眼中有 10^7 个成视网膜细胞，若所有细胞都携有癌基因就使得产生眼瘤的概率相对较高。而散发型是非遗传形式，由于需要两个突变均在同一细胞发生，这种突变的概率很低，所以通常仅有一只眼发病(图 S3.2)。值得注意的是虽然遗传性成视膜细胞瘤只占病例中的少部分，但它却占肿瘤的多数。成视网细胞瘤的"双击"假说（"two-hit"）也可从杂合性的丢失的证据得到证明。对家族疾病的遗传分析暂定成视网膜细胞瘤基因（*RB1*）位于人 13 号染色体。用与 *RB1* 基因紧密连锁的标记做探针进行杂交显示连锁序列杂合病人的瘤细胞仅有该序列的单拷贝，正说明瘤细胞中所推测的 *RB1* 基因所在区有缺失，而正常细胞中并没有缺失。

***RB1* 基因**

在鉴定出由最紧密连锁的标记序列（总是与 *RB1* 基因一起遗传的特定染色体 DNA 序列）所限定的 13 号染色体上特定区域的 DNA 序列后分离出 *RB1* 基因。*RB1* 编码一个 110kDa 的可与 DNA 结合的磷蛋白，具有抑制原癌基因诸如 *myc*、*fos* 转录的作用，在成视网膜细胞瘤中 *RB1* mRNA不是没有就是不正常，将已克隆的、正常的 *RB1* 基因转染培养状

态下的瘤细胞，结果瘤细胞转变成正常状态，这就清楚地证实了 *RB1* 基因的功能。出人意料的是在乳腺癌、结肠癌和肺癌患者中也检测到了 *RB1* 的突变型。

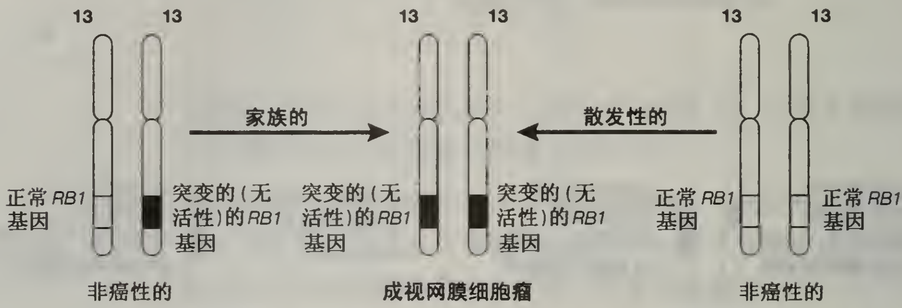


图 S3.2 成视网膜细胞瘤是 13 号染色体上 *RB1* 基因的两个拷贝均失活的结果。由于正常基因两个拷贝的突变（散发性），或者一个失活拷贝的遗传再加上另一个维持功能拷贝的获得突变（家族性）均可发病

***P53* 基因**

同样的技术被应用来鉴定或分离其他癌症中相关的肿瘤抑制基因，其中真正具肿瘤抑制作用的基因被称为 *P53*。*P53* 基因位于 17 号染色体的短臂上，该区域的缺失与近 50％的人癌相关。*P53* 的 mRNA 长 2.2～2.5kb，编码一个 52kDa 的核蛋白。该蛋白在大多数细胞内浓度很低，且半衰期很短（6～20min）。令人迷惑的是 *P53* 同时具有癌基因和肿瘤抑制基因的某些特性：

- 已发现多种突变（点突变、缺失、插入等）都能使 *P53* 转变成致癌性基因。*P53* 的突变体与 *ras* 癌基因一起共转染能使正常鼠成纤维细胞转型。在癌细胞里，P53 蛋白半衰期变长(4～8h)，结果是蛋白浓度的提高，所有这些都表明 *P53* 是一个癌基因。
- 在许多肿瘤中也观察到 17 号染色体的短臂上有相同缺失。在脑癌、乳腺癌、肺癌和结肠癌中，凡 *P53* 基因缺失，另一个等位基因必然也已发生突变，这表明 *P53* 也是一个肿瘤抑制基因。

看来较好的解释只能是 P53 以二聚体形式作用。当突变了的（失活的）P53 蛋白存在时，它与野生型蛋白结合形成一个失活的二聚体复合体（图 S3.3）。这就是所谓的**显性负**（dominant-negative）效应。然而由于正常蛋白二聚体的形成，由突变基因所引起的正常 *P53* 基因的失活不可能达到 100％，而另一个正常 *P53* 基因的丢失就相当于完全逃脱了该基因的肿瘤抑制作用。

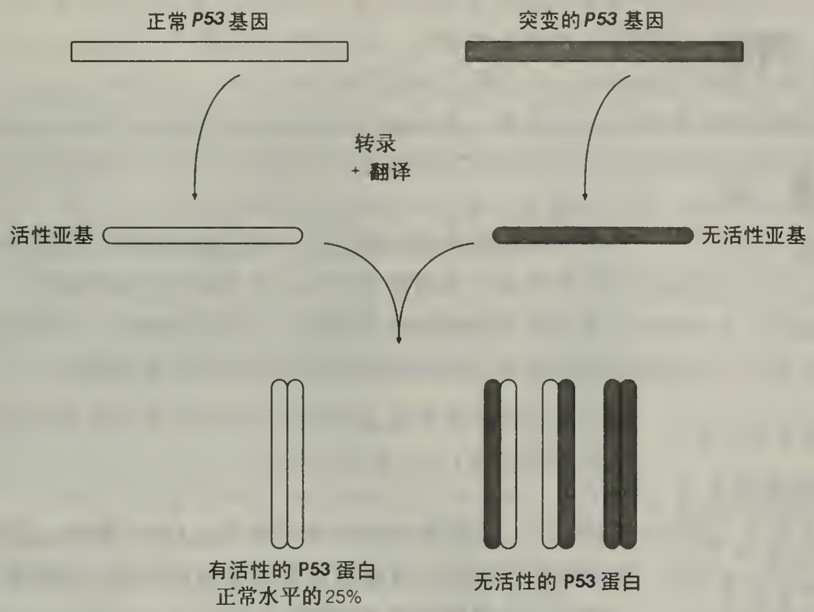


图 S3.3 突变了的 *P53* 基因的显性失活效应取决于突变蛋白与正常蛋白形成二聚体的程度，并使正常蛋白失活

（陆 韵 译 刘进元 校）

S4 凋亡

要 点

凋 亡	在多细胞生物体中，凋亡是一类引起细胞死亡的重要途径。凋亡的发生经历一系列特定过程，并受保守机制的调控。凋亡的根本作用是在发育过程中去除那些不必要的细胞。细胞的分裂与凋亡间的平衡对于维持细胞数量的稳定是非常重要的。
损伤或危险性细胞的清除	凋亡在清除损伤细胞或危险性细胞方面发挥重要作用，例如自身免疫的预防或 DNA 损伤的应答。
凋亡过程中细胞变化	在凋亡中，细胞核内的染色质浓缩、DNA 断裂。发生凋亡的细胞脱离周围细胞，皱缩并且断裂成凋亡小体。邻近细胞识别凋亡小体，并通过胞吞的方式清除他们。
线虫中的凋亡作用	线虫（*Caenorhabditis elegans*）的成虫由 959 个固定数量的细胞构成，是从早期形成的 1090 个细胞，经发育过程中刚好 131 个细胞凋亡的结果。*ced* -*3* 和 *ced* -*4* 基因为细胞凋亡所需，通过抑制 *ced* -*9* 基因的产物，引起大量凋亡。
哺乳动物中的凋亡	哺乳动物中与线虫 *ced* -*9* 基因相似的是 *bcl* -*2* 基因，其作用是抑制凋亡。*bcl* -*2* 的其他类似物在维持细胞存活和死亡平衡中，或者抑制凋亡，或者增强凋亡。哺乳动物中的 *ced* -*3* 基因类似物编码一种称为 caspase 蛋白酶，在凋亡过程中起重要作用。
凋亡在疾病和癌症中的作用	凋亡的失调在疾病和癌症中很关键。一些原癌基因例如 *bcl* -*2* 阻碍凋亡发生，反映了在肿瘤形成中抑制凋亡的作用。c-*myc* 原癌基因具有双重作用既可以促进细胞增殖，同时在某些生长信号缺乏时诱发凋亡。癌症化疗一般使用可引发凋亡的 DNA 损伤药物。

相关主题

细胞周期（E3）
诱变（F1）
DNA 损伤（F2）
DNA 修复（F3）
限制酶与电泳（G3）
癌基因的分类（S2）
肿瘤抑制基因（S3）

凋亡

凋亡是一种细胞的正常死亡机制，其过程涉及一整套程序化的生化和形态的变化。在多细胞生物中凋亡是一种频繁而又普遍的过程，并在正常组织的形成、维持和成型过程中起重要作用。凋亡发生在发育过程的几乎

所有组织以及许多成体组织。例如凋亡在手指间缝隙的形成过程中起重要作用，否则人手则会成为网状物。在两栖动物变态过程中，蝌蚪尾巴的丢失也是一种凋亡。在许多细胞类型中，凋亡是一种内在自我破坏的途径。当生长信号缺乏时该途径会被自动引发，而生长信号则是促使细胞存活。这种细胞分裂与凋亡的平衡对于维持生物体内细胞数量的稳定是至关重要的。所以在成体中组织体积的保持不仅通过细胞的增殖、分化和迁移，而且在很大程度上通过凋亡来控制细胞的消除。

凋亡不是细胞死亡的惟一途径。在坏死过程中，细胞膜失去完整性，细胞裂解并释放内容物。通常不发生完整生物体内容物的释放。因此凋亡是生理性细胞死亡的主要途径，也可称之为**程序性细胞死亡**（也被认为是细胞自杀）。

损伤或危险性细胞的清除

凋亡负责清除掉损伤细胞或危险性细胞。胸腺中大于 90% 的免疫细胞都经凋亡而除去。该过程对清除自身反应性 T 淋巴细胞十分重要，否则会使免疫系统攻击宿主细胞而引起自身免疫病。当 T 细胞杀死其他细胞时，是通过激活凋亡途径并诱导细胞自杀而实现的。许多感染病毒的细胞通过凋亡而限制病毒在机体内的传播。由于 DNA 过度损伤而导致细胞无法修复时，肿瘤抑制子 *P53* 有双重作用，首先抑制细胞周期（参见 E3）然后诱导凋亡。

凋亡过程中的细胞变化

凋亡过程中细胞核皱缩、染色体浓缩。在这个过程发生时，DNA 在核小体之间被核酸酶酶解切成碎片，通过 DNA 的凝胶电泳可以得到展示，（图 S4.1a，参见 G3）凋亡的细胞脱离周围细胞，变圆、收缩、片状化形成凋亡小体（图 S4.1b）通常在凋亡小体内有完整的细胞器和原生质膜。邻近细胞快速吞噬并破坏凋亡小体。目前认为凋亡小体细胞表面的变化在指导胞吞过程中起重要作用。凋亡发生半小时后可见形态学特点的变化。

线虫中的凋亡作用

对于线虫（*C. elegans*），每个雌雄同体的成虫都是相同的，由 959 个细胞构成。线虫的细胞谱系是被严格调控的，在幼虫发育过程中，早期形成 1090 个细胞，其中的 131 个细胞通个凋亡而被除去。*ced -3* 和 *ced -4* 两种基因的产生是发育过程中凋亡发生所必需的。如果其中任何一个基因由于突变而没有被激活，那么这 131 个细胞就没有一个会死亡。另一个 *ced -9* 基因所编码的蛋白则执行相反的功能即抑制凋亡。突变造成 *ced -9* 基因失活会引起细胞过量死亡，甚至在正常发育过程中不该死亡的细胞也会死亡。因此 *ced -9* 为那些在正常情况下不死亡细胞的存活所需，并抑制细胞死亡的整体进程。

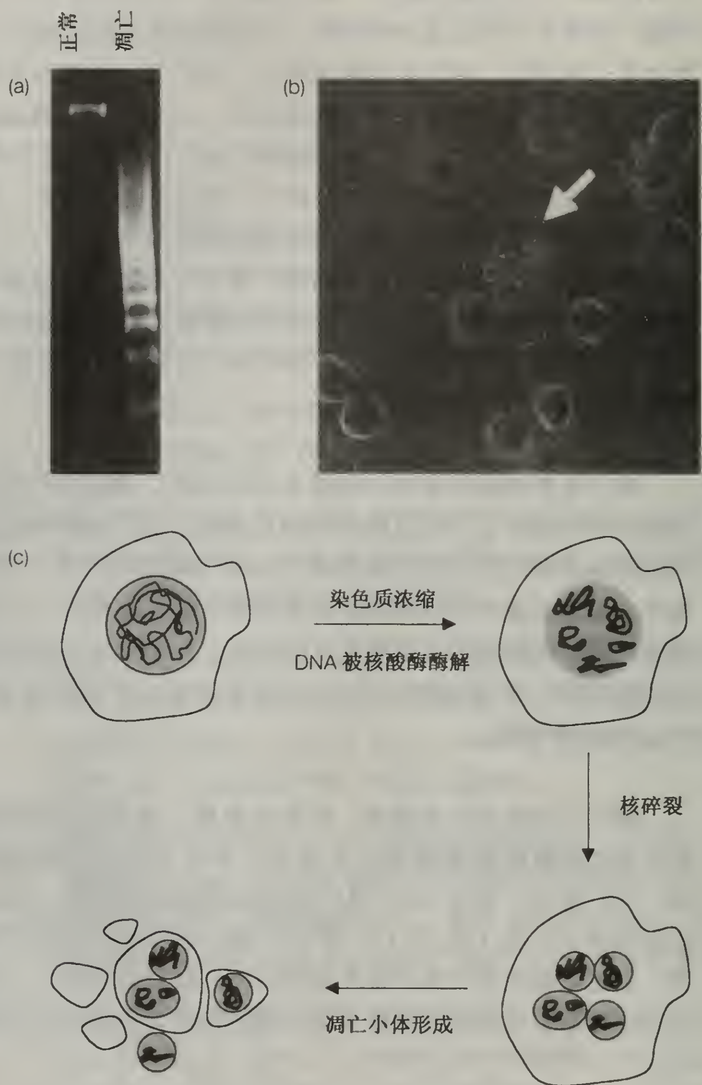


图 S4.1　凋亡。(a) 处于凋亡细胞中 DNA 的特征性梯状带；(b) 显示凋亡小体 T-细胞的显微图；(c) 凋亡过程各阶段的面图

哺乳动物中的凋亡

哺乳动物的原癌基因 *bcl*-*2* 与线虫的凋亡抑制基因 *ced*-*9* 同源，*bcl*-*2*基因是被发现的一类具有新功能的原癌基因（参见 S2）中的第一个基因。这一家族的成员主要是抑制细胞凋亡而不是促进细胞增殖。现在知道在哺乳动物体中存在一套抑制细胞死亡的基因，其中的一些是 *bcl*-*2* 的同源物。包括 *bcl*-*2* 类似物的另一组蛋白如 *bax* 具有促进细胞死亡作用。BCL-2 和 BAX 蛋白在细胞内相互结合，因此细胞内信号转导通路所调节的凋亡有可能部分是通过改变其抑制死亡和促进死亡蛋白在细胞内的相对水平来实现的。线虫 *ced*-*3*、杀手基因的哺乳动物同源物也已被鉴

定。*ced* -*3* 基因编码的多肽与半胱氨酸蛋白酶家族同源，其中代表性的原型是白介素-1 转换酶（ICE）。这些白介素-1 转换酶也被称之为 caspase 酶，他们主要执行凋亡作用。由此可见细胞凋亡不仅是最重要的生命过程而且在进化上是保守的。

凋亡在疾病和癌症中的作用

细胞凋亡的失控会导致许多疾病，包括神经退行性病变、免疫缺陷、心脏病发作或中风后的细胞死亡以及病毒或细菌感染。尤为重要的凋亡的丧失会引起癌症。癌症是多细胞生物体的一种疾病，主要是由于细胞增殖（参见 E3）和细胞死亡之间的平衡被打破。许多原癌基因（参见 S2）调节细胞分裂，但另一些则调节凋亡（如 *bcl* -*2*），反映了平衡在这些过程中的重要性。原癌基因 c-*myc*（参见 S2）具有双重作用，既可以激活细胞分裂又可以引发凋亡，当生长因子缺乏或细胞处于 DNA 损伤时，c-*myc* 起凋亡作用。因此由于基因突变造成的凋亡途径的丧失或相当低水平的调控结果会导致癌症。另一方面某些基因如 *bcl* -*2* 这样抑制正常的凋亡途径基因的过度表达也会引起癌症。许多癌症治疗采用 DNA 损伤药物（原则参见 F2），这些药物可以杀死分裂中的细胞。目前已认识到抗癌药与其说没有特异性作用倒不如说是通过诱发凋亡而杀死癌细胞。对这些药物产生抗药性的主要机制之一在于凋亡途径的抑制。在癌症患者中，有 50% 是由于肿瘤抑制基因 *P53* 的突变造成的（参见 S3）。其结果 *P53* 被称为“基因组的守护者”。在对 DNA 损伤的反应中，*P53* 在引起凋亡过程中起中心作用，同时它又启动 DNA 修复过程（参见 F3）且抑制细胞周期行进（参见 E3）。当损伤太大无法修复时，诱导凋亡不仅可清除癌细胞，而且可抑制突变细胞在细胞群内的增殖，否则会因细胞变化而生成更加恶性的肿瘤。

（冯　海　译　刘进元　校）

进一步阅读文献

现今已有大量分子生物学和生物化学领域的综合型的教科书,但无一能满足读者的全部需要。不同的读者会根据自身的爱好选择不同的教科书,因此我们认为向读者推荐说某书胜过另一本书的做法不是非常确当。相反在此列出了一些根据我们经验认为对读者非常有用的论著。

一般读物

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1994) *Molecular Biology of the Cell*, 3rd Edn. Garland Publishing, New York.

Brown, T.A. (1999) *Genomes*. BIOS Scientific Publishers, Oxford.

Freifelder, D.M. (1987) *Molecular Biology*, 2nd Edn. Jones and Bartlett, Boston, MA.

Lewin, B. (2000) Genes VII. Oxford University Press, Oxford.

Lodish, H., Baltimore, D., Berk, A., Zipursky, S.L., Matsudaira, P. and Darnell, J. (2000) *Molecular Cell Biology*, 4th Edn. Scientific American Books, W.H. Freeman, New York.

Twyman, R.M. (1998) *Advanced Molecular Biology*. BIOS Scientific Publishers, Oxford.

Voet, D. and Voet, J.G. (1995) *Biochemistry*, 2nd Edn. John Wiley & Sons, New York.

扩展读物

以下所选的文章是推荐给那些希望更深入地了解某些主题的读者的。这些文章对于大多数大学一年级学生而言太过高深,但却是他们日后可能会研究的主题的非常有用的信息源。

A

Bretscher, M.S. (1985) The molecules of the cell membrane. *Sci. Amer*. **253**, 86–90.

de Duve, C. (1996) The birth of complex cells. *Sci. Amer*. **274** (4), 38~45.

Gupta, R.S. and Golding, G.B. (1996) The origin of the eukaryotic cell. *Trends Biochem. Sci*. **21**, 166~171.

Pumplin, D.W. and Bloch, R.J. (1993) The membrane skeleton. *Trends Cell Biol*. **3**, 113~117.

Rothman, J.E. and Orci, L. (1996) Budding vesicles in living cells. *Sci. Amer*. **274** (3) 70~75.

B

Byard, E.H. and Lange, B.M.H. (1991) Tubulin and microtubules. *Essays Biochem*. **26**, 13~25.

Csermely, P. (1997) Proteins, RNAs and chaperones in enzyme evolution: a folding perspective. *Trends Biochem. Sci*. **22**, 147~149.

Darby, N.J. and Creighton, T.E. (1993) *Protein Structure: In Focus*. IRL Press, Oxford.

Doolittle, R.F. (1985) Proteins. *Sci. Amer.* **253** (4) 74~81.

Gahmberg, C.G. and Tolvanen, M. (1996) Why mammalian cell surface proteins are glycoproteins. *Trends Biochem. Sci.* **21**, 308~311.

Richards, F.M. (1991) The protein folding problem. *Sci. Amer.* **264** (1), 34~41.

C

Bates, A.D. and Maxwell, A. (1993) *DNA Topology: In Focus.* IRL Press, Oxford.

Drlica, K. (1990) Bacterial topoisomerases and the control of DNA supercoiling. *Trends Genet.* **6** (12), 433~437.

Maxwell, A. (1997) DNA gyrase as a drug target. *Trends Microbiol.* **5** (3), 102~109.

Neidle, S. (1994) *DNA Structure and Recognition: In Focus.* IRL Press, Oxford.

Roca, J. (1995) The mechanisms of DNA topoisomerases. *Trends Biochem. Sci.* **20** (4), 156~160.

D

Grosschedl, R., Giese, K. and Pagel, J. (1994) HMG domain proteins – architectural elements in the assembly of nucleoprotein structures. *Trends Genet.* **10** (3), 94~100.

Jones, P.A., Rideout, W.M., Shen, J.C., Spruck, C.H. and Tsai, Y.C. (1992) Methylation, mutation and cancer. *Bioessays* **14** (1), 33~36.

Roth, S.Y. and Allis, C.D. (1992) Chromatin condensation - does histone H1 dephosphorylation play a role? *Trends Biochem. Sci.* **17** (3), 93~98.

Travers, A.A. (1994) Chromatin structure and dynamics. *Bioessays* **16** (9), 657~662.

Wade, P.A., Pruss, D. and Wolffe, A.P. (1997) Histone acetylation: chromatin in action. *Trends Biochem. Sci.* **22** (4), 128~132.

E

Adams, R.L.P. (1991) *DNA Replication: In Focus.* IRL Press, Oxford.

Chong, J.P.J., Thommes, P. and Blow, J.J. (1996) The role of MCM/P1 proteins in the licensing of DNA replication. *Trends Biochem. Sci.* **21**, 102~106.

Diller, J.D. and Raghuraman, M.K. (1994) Eukaryotic replication origins - control in space and time. *Trends Biochem. Sci.* **19**, 320~325.

Greider, C.W. and Blackburn, E.H. (1996) Telomeres, telomerases and cancer. *Sci. Amer.* **274** (2) 80~85.

Hamlin, J.L. (1992) Mammalian origins of DNA replication. *Bioessays* **14**, 651~660.

Hozak, P. and Cook, P.R. (1994) Replication factories. *Trends Cell Biol.* **4**, 48~52.

Kornberg, A. and Baker, T. (1991) *DNA Replication*, 2nd Edn. W.H. Freeman, New York.

Murray, A.W. and Kirschner, M.W. (1991) What controls the cell cycle? *Sci. Amer.* **264** (3), 56~63.

F

Bridges, B.A. (1997) DNA turnover and mutation in resting cells. *Bioessays* **19**, 347~52.

Friedberg, E.C., Walker, G.C. and Siede, W. (1995) *DNA Repair and Mutagenesis* American Society for Microbiology, Washington, DC.

Shinagawa, H. and Iwasaki, H. (1996) Processing the Holliday junction in homologous recombination. *Trends Biochem. Sci.* **21**, 107~111.

Tanaka, K. and Wood, R.D. (1994) *Xeroderma pigmentosum* and nucleotide excision repair of DNA. *Trends Biochem. Sci.* **19**, 83~86.

Trends in Biochemical Sciences, Vol. 20, No. 10, 1995. Whole issue devoted to articles on DNA repair.

G

Brown T.A. (1995) *Gene Cloning: An Introduction*, 3rd Edn. Chapman and Hall, London.

Watson, J.D., Gilman, M., Witkowski, J. and Zoller, M. (1992) *Recombinant DNA*, 2nd Edn. W.H. Freeman, New York.

H Anand, R. (1992) Yeast artificial chromosomes (YACs) and the analysis of complex genomes. *Trends Biotechnol.* **10** (1–2), 35~40.

Calos, M.P. (1996) The potential of extrachromosomal replicating vectors for gene-therapy. *Trends Genet.* **12** (11), 463~466.

Conzelmann, K.K. and Meyers, G. (1996) Genetic-engineering of animal RNA viruses. *Trends Microbiol.* **4** (10), 386~393.

Tinland, B. (1996) The integration of T-DNA into plant genomes. *Trends Plant Sci.* **1** (6), 178~184.

Walden, R. and Wingender, R. (1995) Gene-transfer and plant-regeneration techniques. *Trends Biotechnol.* **13** (9), 324~331.

I Brown T.A. (1995) *Gene Cloning: An Introduction*, 3rd Edn. Chapman and Hall, London.

Watson, J.D., Gilman, M., Witkowski, J. and Zoller, M. (1992) *Recombinant DNA*, 2nd Edn. W.H. Freeman, New York.

J Adams, R.L.P., Knowler, J.T. and Leader, D.P. (1992) *The Biochemistry of the Nucleic Acids*, 11th Edn. Chapman and Hall, London.

French Anderson, W. (1995) Gene therapy. *Sci. Amer.* **273**, 96~99.

Gurdon, J.B. and Colman, A. (1999) The future of cloning. *Nature* **402**, 743~746.

Heiter, P. (1999) What do yeast proteins do? *Nature* **402**, 362~363.

Mullis, K.B. (1990) The unusual origins of the polymerase chain reaction. *Sci. Amer.* **262**, 36~41.

K Beebee, T.J.C. and Burke, J. (1992) *Gene Structure and Transcription: In Focus.* IRL Press, Oxford.

Ptashne, M. (1986) *A Genetic Switch.* Cell Press, Palo Alto, CA.

L Beebee, T.J.C. and Burke, J. (1992) *Gene Structure and Transcription: In Focus.* IRL Press, Oxford.

Pugh, B.F. (1996) Mechanisms of transcription complex assembly. *Curr. Opin. Cell Biol.* **8**, 303~311.

M Burley, S.K. (1996) The TATA box-binding protein. *Curr. Opin. Struct. Biol.* **6**, 69~75.

Chalut, C., Moncollin, V. and Egly, J.M. (1994) Transcription by RNA polymerase II. *Bioessays* **16**, 651~655.

Geiduschek, E.P. and Kassavetis, G.A. (1995) Comparing transcriptional initiation by RNA polymerase I and RNA polymerase III. *Curr. Opin. Cell Biol.* **7**, 344~351.

Roeder, R.G. (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**, 327~334.

Verrijzer, C.P. and Tijan, R. (1996) TAFs mediate transcriptional activation and promoter selectivity. *Trends Biochem. Sci.* **21**, 338~342.

N Laemmli, U.K. and Tijan, R. (1996) Nucleus and gene expression – a nuclear traffic jam – unraveling multicomponent machines and compartments. *Curr. Opin. Cell Biol.* 8, 299~302.

Maldonado, E. and Reinberg, D. (1995) News on initiation and elongation of transcription by RNA polymerase II. *Curr. Opin. Cell Biol.* **7**, 352~361.
Tijan, R. (1995) Molecular machines that control genes. *Sci. Amer.* **272**, 38~45.
Wolberger, C. (1996) Homeodomain interactions. *Curr. Opin. Struct. Biol.* **6**, 62~68.

O Adams, M.D., Rudner, D.Z. and Rio, D.C. (1996) Biochemistry and regulation of pre-mRNA splicing. *Curr. Opin. Cell Biol.* **8**, 331~339.
Morrissey, J.P. and Tollervey, D. (1995) Birth of the snoRNPs - the evolution of RNase MRP and and the eukaryotic pre-rRNA-processing system. *Trends Biochem. Sci.* **20**, 78~82.
Nagai, K. (1996) RNA–protein complexes. *Curr. Opin. Struct. Biol.* **6**, 53–61.
Ross, J. (1996) Control of mRNA stability in higher eukaryotes. *Trends Genet.* **12**, 171~175.
Scott, W.G. and Klug, A. (1996) Ribozymes: structure and mechanism of RNA catalysis. *Trends Biochem. Sci.* **21**, 220~224.
Smith, H.C. and Snowden, M.P. (1996) Base modification mRNA editing through deamination - the good, the bad and the unregulated. *Trends Genet.* **12**, 418~424.
Tuschl, T., Thomson, J.B. and Eckstein, F. (1995) RNA cleavage by small catalytic RNAs. *Curr. Opin. Struct. Biol.* **5**, 296~302.
Wahle, E. and Keller, W. (1996) The biochemistry of polyadenylation. *Trends Biochem. Sci.* **21**, 247~250.

P Arnstein, H.R.V. and Cox, R.A. (1992) *Protein Biosynthesis: In Focus.* IRL Press, Oxford.
Moras, D. (1992) Structural and functional relationships between aminoacyl-tRNA synthetases. *Trends Biochem. Sci.* **17**, 159~164.

Q Garrett, R. (1999) Mechanism of the ribosome. *Nature* **400**, 811~812.
Liu, R. and Neupert, W. (1996) Mechanisms of protein import across the outer mitochondrial membrane. *Trends Cell Biol.* **6**, 56~61.
Martoglio, B. and Dobberstein, B. (1996) Snapshots of membrane-translocating proteins. *Trends Cell Biol.* **6**, 142~147.
Rassow, J. and Pfanner, N. (1996) Protein biogenesis – chaperones for nascent polypeptides, *Curr. Biol.* **6**, 115~118.
Riis, B., Rattan, S.I.S., Clark, B.F.C. and Merrick, W.C. (1990) Eukaryotic protein elongation factors. *Trends Biochem. Sci.* **15**, 420~424.

R Cann, A.J. (1993) *Principles of Molecular Virology*, Academic Press, London.
Liljas, I. (1996) Viruses, *Curr. Opin. Struct. Biol.* **6**, 151~156.

S Harris, C.C. (1993) p53: at the crossroads of molecular carcinogenesis and risk assessment. *Science* **26**, 1980~1981.
Macdonald, F. and Ford, C.H.J. (1991) *Oncogenes and Tumor Suppressor Genes.* BIOS Scientific Publishers, Oxford.
Weinberg, R.A. (1989) Oncogenes, antioncogenes and the molecular bases of multistep carcinogenesis. *Cancer Res.* **49**, 3713~3721.

选 择 题

A 细胞与大分子

1. 分泌蛋白的糖基化作用发生在______中。
   A. 线粒体
   B. 过氧化物酶体
   C. 内质网
   D. 核
2. 以下哪一个是核蛋白?
   A. 角蛋白
   B. 染色质
   C. 组蛋白
   D. 蛋白聚糖
3. 以下哪一个不是多糖?
   A. 几丁质
   B. 支链淀粉
   C. 黏多糖
   D. 甘油
4. 跨膜蛋白______。
   A. 将两个脂双层连接一起
   B. 有胞内和胞外结构域
   C. 完全包埋在膜内
   D. 很容易从膜上去除下来

B 蛋白质结构

1. 以下哪一个是亚氨基酸?
   A. 脯氨酸
   B. 羟赖氨酸
   C. 色氨酸
   D. 组氨酸
2. 在不同物种中执行同一生化功能的蛋白家族成员叫做______。
   A. 共生同源物
   B. 结构类似物
   C. 异源物
   D. 直向同源物

3. 以下哪一个不是蛋白质二级结构?
   A. α-螺旋
   B. 三股螺旋
   C. 双螺旋
   D. β-折叠片
4. 在等电聚焦电泳中,蛋白质按______分离。
   A. pH 梯度
   B. 盐梯度
   C. 密度梯度
   D. 温度梯度
5. Edman 降解法______对多肽进行序列测定。
   A. 用一段 cDNA 序列
   B. 按照质量
   C. 从 C 端到 N 端
   D. 从 N 端到 C 端

C 核酸的性质

1. DNA 中的一段 5′-AGTCTGACT-3′序列等同于 RNA 中的哪一段?
   A. 5′-AGUCUGACU-3′
   B. 5′-UGTCTGUTC-3′
   C. 5′-UCAGUCUGA-3′
   D. 5′-AGUCAGACU-3′
2. 以下哪一个是对 A-DNA 的正确描述?
   A. 是右手反向平行双螺旋,每圈 10 个碱基,碱基平面与螺旋轴垂直
   B. 是左手反向平行双螺旋,每圈 12 个碱基,由交替的嘌呤-嘧啶序列形成
   C. 是右手反向平行双螺旋,每圈 11 个碱基,碱基的朝向与螺旋轴有关
   D. 是球状结构,由单链核酸中形成的短的分子内螺旋构成
3. 双链 DNA 的变性涉及______。
   A. 断裂为短的双链片段
   B. 分开成为单链
   C. DNA 骨架的分解
   D. 碱基从糖-磷酸骨架上分裂下来
4. 以下哪一个在 260nm 波长下具有最大的单位质量吸收?
   A. 双链 DNA
   B. 单核苷酸
   C. RNA
   D. 蛋白质
5. I 型 DNA 拓朴异构酶______。
   A. 使连接数改变 ±2

B. 需要 ATP
C. 将 DNA 双螺旋中的一条链打断
D. 是抗生素类药物的靶向

D 原核与真核生物的染色体结构

1. 以下哪一个是大肠杆菌和真核生物染色体的共同之处?
A. DNA 是环状的
B. DNA 被包装成核小体
C. DNA 位于核中
D. DNA 是负超螺旋
2. 由 166pb DNA、组蛋白八聚体和组蛋白 H1 组成的复合体叫做______。
A. 核小体核心
B. 螺线管
C. 30nm 纤维
D. 核小体
3. 分裂间期染色体转录发生在什么区域?
A. 端粒
B. 着丝粒
C. 常染色质
D. 异染色质
4. 以下关于 CpG 岛和甲基化的陈述哪一个是错误的?
A. CpG 岛对 DNase I 有部分抗性
B. CpG 甲基化是真核生物对 CpG 突变为 TpG 的反应
C. CpG 岛出现在活性基因启动子附近
D. CpG 甲基化伴随着染色质失活
5. 以下哪一个是高度重复 DNA?
A. *Alu* 元件
B. 组蛋白基因簇
C. DNA 微卫星
D. 散在重复 DNA

E DNA 复制

1. 一个典型哺乳动物细胞中复制子的数目是______。
A. 40～200
B. 400
C. 1000～2000
D. 50 000～100 000
2. 原核生物中,后随链的引物是被______清除的。
A. 3′至 5′外切核酸酶

B. DNA 连接酶

C. DNA 聚合酶 I

D. DNA 聚合酶Ⅲ

3. 在大肠杆菌复制起始点处的基本起始蛋白是______。

A. DnaA

B. DnaB

C. DnaC

D. DnaE

4. 如果细胞在 R 点前缺乏有丝分裂剂,它将进入______期。

A. G_1

B. S

C. G_2

D. G_0

5. 以下哪一个陈述是正确的?

A. 一旦细胞越过 R 点,细胞分裂就不可避免

B. G_1 细胞周期蛋白 - CDK 复合体对 RB 的磷酸化是进入 S 期的关键

C. G_1 细胞周期蛋白 - CDK 复合体对 E2F 的磷酸化是进入 S 期的关键

D. 细胞周期蛋白 D1 和 INK4 P16 是肿瘤抑制蛋白

6. 在真核生物中,常染色质优先在______复制。

A. S 期早期

B. S 期中期

C. S 期晚期

D. G_2 期

7. 原核生物的质粒如果含有酵母的______就可以在酵母细胞中复制。

A. ORC

B. CDK

C. ARS

D. RNA

F DNA 损伤、修复与重组

1. 每掺入一个核苷酸,大肠杆菌的自发突变频率是______。

A. $1/10^6$

B. $1/10^8$

C. $1/10^9$

D. $1/10^{10}$

2. 羟自由基对 DNA 的作用产生了大量______。

A. 嘧啶二聚体

B. 8-氧化鸟嘌呤

C. O^6-甲基鸟嘌呤

D. 7-羟甲基鸟嘌呤

3. 在大肠杆菌的甲基指向错配修复过程中，含有错配碱基的子链被______切开。

A. MutH 内切核酸酶

B. UvrABC 内切核酸酶

C. AP 内切核酸酶

D. 3′至 5′外切核酸酶

4. 异常重组又被称为______。

A. 位点特异性重组

B. 转座

C. 同源重组

D. 转移损伤 DNA 合成

5. 在患有______的个体中缺乏对紫外线诱导 DNA 损伤的切除修复。

A. 遗传型结肠癌

B. Crohn 氏症

C. 典型的 DNA 修复酶缺乏病

D. DNA 修复酶缺乏病的变型

G 基因操作

1. 细菌培养物中某种质粒的存在通常用______来确定。

A. 蓝-白颜色筛选

B. 在抗生素存在下生长

C. 一种限制酶降解

D. 琼脂糖凝胶电泳

2. 碱性磷酸酶______。

A. 将双链 DNA 片段连接起来

B. 在 DNA 的 5′端加上一个磷酸基团

C. 切断双链 DNA

D. 从 DNA 的 5′端去除一个磷酸基团

3. 转化是______。

A. 细菌吸收一个质粒

B. 细菌表达一个基因

C. 细菌吸收一个噬菌体

D. 从细菌中分离一个质粒

4. T4 DNA 连接酶______。

A. 需要 ATP

B. 将具有相邻 3′-磷酸和 5′-羟基的双链 DNA 片段连接起来

C. 需要 NADH

D. 将单链 DNA 连接起来

5. 在琼脂糖凝胶电泳中______。

A. DNA 向负极移动
B. 超螺旋质粒泳动得比其带切口的异构体慢
C. 大分子比小分子泳动快
D. 溴化乙锭被用来显色 DNA

H 克隆载体

1. 蓝-白筛选被用于______。
   A. 检测细菌中某一质粒的存在
   B. 提示一个克隆化 DNA 片段的性质
   C. 表达一个克隆化基因
   D. 检测质粒中某一插入片段的存在
2. 多克隆位点______。
   A. 含有许多拷贝的克隆化基因
   B. 使得对克隆所用的限制酶的选择具有灵活性
   C. 使得对克隆所用的生物种类的选择具有灵活性
   D. 含有许多拷贝的同一种限制酶切位点
3. λ 噬菌体对大肠杆菌的感染一般用______检测。
   A. 细菌对某种抗生素的抗性
   B. 单菌落在琼脂平板上的生长
   C. 琼脂平板上产生裂解细菌区域
   D. 细菌 DNA 的限制性消化
4. 哪一个载体适用于 150kb DNA 片段的克隆?
   A. 质粒
   B. λ 载体
   C. BAC
   D. YAC
5. 要在植物中表达一个外源基因,你会选用哪个载体?
   A. 杆状病毒载体
   B. 反转录病毒载体
   C. Yep 载体
   D. T-DNA 载体

I 基因文库与筛选

1. 以下关于基因组文库陈述哪两个是错误的?
   A. 基因组文库是用 cDNA 制备的
   B. 包含某种生物所有基因的基因组文库,必定是代表性文库
   C. 若要含有某种生物的所有基因,基因组文库就必须含有最少的重组子
   D. DNA 必须先被切成适合于所用的载体的大小
   E. 真核生物 DNA 的基因组文库通常用质粒载体

2. 以下对 cDNA 克隆步骤的描述哪一个是正确的？
   A. mRNA 的反转录，第二链合成，cDNA 末端修饰，连接到载体上
   B. mRNA 制备，用反转录酶合成 cDNA，用末端转移酶合成第二链，连接到载体上
   C. 用 RNA 聚合酶合成 mRNA，mRNA 的反转录，第二链合成，连接到载体上
   D. 双链 cDNA 合成，限制酶消化，加接头，连接到载体上
3. 以下哪一种方法对于筛选文库是无效的？
   A. 用核酸探针对菌落和(或)噬菌体溶出的 DNA 进行杂交
   B. 用目标蛋白的抗体筛选表达文库
   C. 从具生物活性的表达文库中筛选克隆
   D. 用抗体作探针对菌落和(或)噬菌体溶出的 DNA 进行杂交

J 克隆 DNA 的分析与应用

1. 一个线性 DNA 片段的一端被 100% 标记且有 3 个 *Eco*R Ⅰ 位点。如果该片段被 *Eco*R Ⅰ 部分消化并得到所有可能产生的片段，将会多少片段被标记，多少片段未被标记？
   A. 4 个标记，6 个未被标记
   B. 4 个标记，4 个未被标记
   C. 3 个标记，5 个未被标记
   D. 3 个标记，3 个未被标记
2. 以下哪一种方法对于标记双链 DNA 是无效的？
   A. 用寡核苷酸激酶进行 5′端标记
   B. 用寡核苷酸激酶进行 3′端标记
   C. 用末端转移酶进行 3′端标记
   D. 用末端转移酶进行 5′端标记
   E. 切口平移
3. 以下关于核酸测序的陈述哪一个是正确的？
   A. Sanger DNA 的测序法包括哌啶的碱基特异性切割
   B. Maxam 和 Gilbert 的 DNA 测序法要用到 DNA 聚合酶和链终止双脱氧核苷酸
   C. RNA 的酶法测序要用到 RNase A、T1、PhyM 和 *B. cereus* RNase
   D. DNA 的酶法测序要用到引物，该引物是用 RNA 聚合酶延伸的
   E. RNA 的酶法测序要用到 RNase T1、U2、PhyM 和 *B. cereus* RNase
4. 以下关于 PCR 的陈述哪一个是错误的？
   A. PCR 循环包括模板变性、引物退火和核苷酸聚合
   B. PCR 要用热稳定的 DNA 聚合物
   C. 理想的 PCR 引物要长度和 G+C 含量都相似
   D. PCR 条件的优化通常包括镁离子浓度的变化和聚合温度的变化
   E. 如果 PCR 的效率达到 100%，那么在几个循环之后目标分子将被扩增 2^n 倍
5. 以下关于基因作图技术的陈述哪两个是正确的？
   A. S1 核酸酶作图法确定的是基因中的非转录区

B. 引物延伸法确定的是转录物的3′端

C. 凝胶阻滞法可以显示蛋白是否能与某一DNA片段结合并阻滞该片段通过琼脂糖凝胶的迁移

D. DNase Ⅰ足迹法确定蛋白结合到某一DNA片段中的哪一个区域

E. 基因启动子中DNA序列的功能可以通过将其连接到报道基因下游并分析其表达来确定

6. 以下关于诱变技术的陈述哪一个是错误的？

A. 外切酶Ⅲ将DNA的一条链从5′凹端按5′到3′方向切除

B. 外切酶Ⅲ将DNA的一条链从3′凹端按3′到5′方向切除

C. 在PCR中可以用诱变引物来引入碱基变化

D. 可以用诱变引物、单链模板和DNA聚合酶来引入碱基变化

E. 可用限制酶产生缺失突变

7. 以下关于基因克隆的陈述哪一个是错误的？

A. 可以用基因克隆来大量表达重组蛋白

B. DNA足迹法被用于检测与DNA结合的蛋白

C. 被克隆的基因可被用于检测致病基因的携带者

D. 基因治疗是通过将正确的基因导入病人体内以纠正原有的错误

E. 遗传修饰生物已被用来生产临床上有重要用途的蛋白质

K　原核生物的转录

1. 以下关于转录的陈述哪两个是正确的？

A. RNA合成按3′至5′方向进行

B. RNA聚合酶按5′至3′方向沿DNA有义链移动

C. RNA聚合酶按5′至3′方向沿DNA模板链移动

D. 转录所得的RNA与模板链互补

E. RNA聚合酶将核糖核苷酸添加到正在生长的RNA链的5′端

F. RNA聚合酶将脱氧核糖核苷酸添加到正在生长的RNA链的5′端

2. 以下关于大肠杆菌RNA聚合酶的陈述哪一个是错误的？

A. 全酶包括σ因子

B. 核心酶包括σ因子

C. 需要Mg^{2+}才有活性

D. 需要Zn^{2+}才有活性

3. 以下哪个陈述不正确？

A. 大肠杆菌RNA聚合酶有两个α亚基

B. 大肠杆菌RNA聚合酶有一个β亚基

C. 大肠杆菌有一个σ因子

D. 大肠杆菌RNA聚合酶β亚基被利福平抑制

E. 利迪链菌素可抑制转录延伸

F. 发夹结构是聚阴离子，它可与β′亚基结合

4. 以下有关大肠杆菌转录的陈述哪一个是正确的?
   A. −10 序列通常正好位于转录起始位点上游 10bp 处
   B. 起始核苷酸通常是 G
   C. −35 和 −10 序列间的间隔序列是保守的
   D. 转录起始位点后的序列对于转录效率不重要
   E. −35 和 −10 序列的距离对于转录效率非常重要
5. 以下有关大肠杆菌转录的陈述哪一个是正确的?
   A. RNA 聚合酶的核心酶与 DNA 的结合是非特异、不稳定的
   B. σ 因子大大提高了聚合酶对正确启动子位点的亲和力
   C. 几乎所有 RNA 起始位点都由一个嘌呤残基组成,而且 A 比 G 出现频率更高
   D. 所有启动子都被负超螺旋所抑制
   E. 终止子通常是 A-U 发夹结构

L 原核生物转录的调控

1. 以下陈述哪两个是正确的?
   A. 含有 5′-GGATCGATCC-3′序列的双链 DNA 序列是一个回文结构
   B. 含有 5′-GGATCCTAGG-3′ 序列的双链 DNA 序列是一个回文结构
   C. Lac 阻抑蛋白抑制聚合酶与 *lac* 启动子的结合
   D. *lac* 操纵子直接被乳糖诱导
   E. Lac 阻抑蛋白与异乳糖的结合降低了它与 *lac* 启动子的亲和力
   F. IPTG 是 *lac* 启动子的天然诱导剂
2. 以下有关分解代谢调节的操纵子的陈述哪一个是错误的?
   A. cAMP 受体蛋白(CRP)和分解代谢激活蛋白(CAP)是同一蛋白的不同名称
   B. 当细胞中存在葡萄糖时 cAMP 水平下降
   C. CRP 结合到 cAMP 上导致转录被激活
   D. 当 cAMP 缺乏时 CRP 结合到 DNA 上
   E. CRP 能使 DNA 弯曲,导致转录被激活
3. 以下有关色氨酸操纵子的陈述哪一个是正确的?
   A. 色氨酸操纵子的 RNA 产物很稳定
   B. Trp 阻抑蛋白是色氨酸操纵子的产物
   C. Trp 阻抑蛋白和 Lac 阻抑蛋白一样,是相同亚基的四聚体
   D. Trp 阻抑蛋白与色氨酸结合
   E. 色氨酸激活色氨酸操纵子的表达
   F. 色氨酸操纵子只被色氨酸阻抑蛋白所调控
4. 以下对于色氨酸操纵子的弱化作用的陈述哪两个是正确的?
   A. 弱化作用是 rho 依赖性的
   B. 弱化序列的缺失会导致色氨酸启动子转录的基本和激活水平的上升
   C. 弱化子位于色氨酸操纵序列的上游
   D. 弱化作用不需要转录和翻译的密切结合

E. 当色氨酸缺乏时，核糖体在前导肽上两个色氨酸密码子之间的暂停导致弱化作用
F. 当色氨酸缺乏时，一种叫做反终子的发夹结构阻止了终止发夹的形成，导致 *trpE* 基因的转录通读

5. 以下关于 σ 因子的陈述哪两个是错误的？
A. 缺乏 σ 因子亚基的大肠杆菌 RNA 聚合酶核心酶不能从启动子处起始转录
B. 不同的 σ 因子识别不同系列的启动子
C. σ 因子识别 −10 和 −35 启动子元件
D. 大肠杆菌热休克启动子有不同的 −35 和 −10 序列且与 17 种不同的热休克 σ 因子结合
E. *B. subtilis* 的孢子形成受多种 σ 因子调节
F. T7 噬菌体表达它自己的 σ 因子，而不是编码它自己的 RNA 聚合酶

M 真核生物的转录

1. 以下对于真核生物 DNA 聚合酶 Ⅰ、Ⅱ 和 Ⅲ 的陈述哪一个是错误的？
A. RNA 聚合酶 Ⅱ 对 α-鹅膏蕈碱非常敏感
B. RNA 聚合酶 Ⅱ 存在于核质中
C. RNA 聚合酶 Ⅲ 转录 tRNA 基因
D. 真核细胞含有除 RNA 聚合酶 Ⅰ、Ⅱ 和 Ⅲ 之外的其他 RNA 聚合酶
E. 每一种 RNA 聚合酶除含有与大肠杆菌 RNA 聚合酶同源的亚基外，还有每一种聚合酶所特有的其他亚基
F. RNA 聚合酶 Ⅱ 的羧基末端含有一段称为羧基末段区(CTD)的只有 7 个氨基酸的短序列，它可能被磷酸化

2. 以下对于 RNA 聚合酶 Ⅰ 基因的陈述哪两个是正确的？
A. RNA 聚合酶 Ⅰ 为核糖体 RNA 转录该基因
B. 人细胞含有 40 簇 5 个拷贝的 tRNA 基因
C. 18S、5.8S 和 28S rRNA 是分开转录的
D. RNA 聚合酶 Ⅰ 的转录发生在核质中
E. RNA 聚合酶 Ⅰ 的转录发生在胞质中
F. rRNA 基因簇被称为核仁组织区域

3. 以下对于 RNA 聚合酶 Ⅰ 转录的陈述哪一个是错误的？
A. 在 RNA 聚合酶 Ⅰ 启动子中，核心元件是上游调控元件(UCE)下游的 1000 个碱基
B. 上游结合因子(UBF)与 UCE 和 RNA 聚合酶 Ⅰ 启动子核心元件的上游部分结合
C. 选择性因子 SL1 使 UBF-DNA 复合体稳定
D. SL1 包含几个亚基，包括 TATA 结合蛋白 TBP
E. *Acanthamoeba* 的 rRNA 基因启动子中只有一个调控元件

4. 以下对于 RNA 聚合酶 Ⅲ 基因的陈述哪两个是正确的？
A. tRNA 基因的转录调控区位于转录起始位点上游
B. tRNA 基因编码区中的高保守序列也是启动子序列
C. TBP 是 TFⅢC 的亚基之一

D. TFⅢB 自身就是一个序列特异性转录因子
E. 人类基因组内有 2000 个拷贝的单基因簇的 5S rRNA 基因

5. 以下陈述哪一个是正确的？
A. RNA 聚合酶Ⅱ只转录编码蛋白的基因
B. TATA 框对转录效率有影响，但对转录起始位点的定位没有影响
C. TBP 与 TATA 框结合
D. 增强子通常位于转录起始位点的上游 100～200 碱基处

6. 以下对于一般转录因子的陈述哪一个是错误的？
A. TFⅡD 结合 TATA 框
B. TFⅡD 是由 TBP 和 $TAF_{Ⅱ}$ 组成的多蛋白复合体
C. TBP 是 RNA 聚合酶Ⅰ、Ⅱ和Ⅲ转录中的共同因子
D. TFⅡB 使 TFⅡD-DNA 复合体稳定
E. RNA 聚合酶结合后，TFⅡE、TFⅡH 和 TFⅡJ 与转录复合体结合
F. TFⅡH 使 CTD 磷酸化

N 真核生物转录的调控

1. 以下对于转录因子的陈述哪两个是正确的？
A. 螺旋-转角-螺旋结构域是转录激活结构域
B. 转录因子的二聚化发生在碱性结构域
C. 亮氨酸拉链与 DNA 结合
D. 当来自几个转录因子的 DNA 结合区和激活区融合在一起时，得到功能性转录因子是可能的
E. 转录因子的同一结构域即可作为阻抑蛋白也可作为激活结构域

2. 以下对于转录调控的陈述哪两个是错误的？
A. SP1 含有两个激活结构域
B. 类固醇激素通过与细胞表面受体结合来调节转录
C. Stat1α 的磷酸化导致它从胞质向核内转移
D. HIV Tat 调节 RNA 聚合酶Ⅱ的磷酸化和持续合成能力
E. MyoD 蛋白能和许多其他的 HLH 转录因子形成异源二聚体
F. 同源异型框是一种保守的 DNA 结合结构域

O RNA 加工与核糖核蛋白复合体

1. 以下几组术语哪一组能正确是描述大肠杆菌大(50S)亚基的几个部分？
A. 杆状、中心突起、凹谷和裂隙
B. 上三分之一、下三分之一、凹谷和杆状
C. 裂隙、凹谷、杆状和小突起
D. 杆状、多肽出口、凹谷和中心突起

2. 大肠杆菌中哪些酶参与合成成熟的 tRNA？
A. RNase A、D、E 和 F

B. RNase D、E、F 和 H

C. RNase D、E、F 和 P

D. RNase A、D、H 和 P

3. 大多数真核生物前体 mRNA 是通过其末端的哪一种修饰而成熟的?

A. 3′端加帽,5′端切割并聚腺苷酸化

B. 在 5′端加一个 GMP,3′端切割并聚腺苷酸化

C. 在 5′端加一个鸟嘌呤,然后切割并聚腺苷酸化以形成 3′端

D. 在 5′端加一个 GMP,聚腺苷酸化,然后切割以形成 3′端

4. 以下对于大多数真核生物前体 mRNA 的剪接加工的描述哪一个是正确的?

A. 在两步反应中,剪接体将外显子以套索的形式去除,然后将两个内含子连接在一起

B. 剪接要求有保守序列,包括 5′-剪接位点,3′-剪接位点、分枝点和聚嘌呤序列

C. U1 snRNP 最初与 5′-剪接位点结合,U2 与分枝点序列结合,然后 tri-snRNP、U4、U5 和 U6 才可以结合上去

D. 在剪接的第一步中,内含子 3′端的 G 与分枝点序列的 A 残基的 2′-羟基连接起来,形成一个套索

P 遗传密码与 tRNA

1. 以下哪一组是对遗传密码特征正确描述?

A. 三联体,简并性,几乎是通用的,无逗点,不重叠

B. 三联体,通用性,无逗点,简并性,不重叠

C. 重叠,三联体,无逗点,简并性,几乎是通用的

D. 重叠,无逗点,不简并,几乎是通用的,三联体

2. 以下对 tRNA 的陈述哪一个是错误的?

A. 大多数 tRNA 长约 76 个残基,其 74、75、76 位点分别为 CCA

B. 许多 tRNA 含有修饰的核苷酸如假尿苷、二氢尿苷、胸腺嘧啶核糖核苷和肌苷

C. tRNA 有着相同的 L 形的三组结构,其一端有三个能与信使 RNA 分子上的反密码子进行碱基配对的核苷酸

D. tRNA 有着相同的三叶草状二级结构,包括 D 环、T 环和反密码子环三个单链环

3. 以下哪三个陈述是正确的? 氨酰 tRNA 合成酶反应______。

A. 将 AMP 连接到 tRNA 的 3′端

B. 是两步反应

C. 将任何的氨基酸连接到 A76 残基核糖的 2′-羟基

D. 是高度特异性的,因为合成酶使用与 tRNA 中相同的元件来区分它们

E. 将 AMP 连接到氨基酸上以产生中间体

F. 在第二步释放焦磷酸

Q 蛋白质合成

1. 以下关于密码子-反密码子相互作用的陈述哪一个是错误的?

A. 它是反向平行的,能包容非标准碱基对

B. 5′反密码子位置的肌苷能与 3′密码子位置的 A、C 或 U 配对

C. 3′反密码子位置的肌苷能与 5′密码子位置的 A、C 或 U 配对

D. 5′反密码子位置不可能出现 A，因为它被反密码子脱氨酶所修饰

2. 以下关于大肠杆菌蛋白合成起始的陈述哪一个是正确的？

A. 起始 tRNA 结合到 Shine-Dalgarno 序列上

B. 有三个起始因子参与，且 IF2 与 GTP 结合

C. 含有 IF1、IF2、IF3、起始 tRNA 和 mRNA 的中间体称为 30S 起始复合体

D. 50S 亚基的结合使 IF1、IF2、GMP 和焦磷酸被释放

E. 当 70S 起始复合体形成时起始过程完成，该复合体中核糖体的 A 位点是起始 tRNA 而 P 位点是空的

3. 以下关于原核生物蛋白合成的延伸的陈述哪一个是错误的？

A. 延伸可分为三步：肽酰 tRNA 的传递、肽键形成和转位

B. 大核糖体亚基的肽酰 tRNA 中心负责肽键形成

C. 在 EF-Fu-Ts 交换循环中，EF-Fu-GTP 可通过 EF-Ts 替换 GDP 而再生

D. EF-G 又叫转位酶，它需要 GTP

4. 大肠杆菌释放因子 1(RF1)识别以下哪个密码子？

A. 只识别 UAA

B. 只识别 UAG

C. 只识别 UGA

D. UGA 和 UAA

E. UAG 和 UAA

F. UAG 和 UGA

5. 以下关于真核生物蛋白合成的起始的陈述哪两个是正确的？

A. 真核生物用 mRNA 扫描法来定位正确的起始密码子

B. 至少有 9 种真核生物起始因子(eIFs)

C. 真核起始使用 *N*-甲酰甲硫氨酸

D. 80S 起始复合体完成起始过程，其 A 位点含有与起始密码子配对的起始 tRNA

E. 在扫描过程中 ATP 被水解成 AMP 和焦磷酸

F. 当 mRNA 与小亚基结合后，起始 tRNA 再结合

6. 以下哪一对蛋白合成因子在原核和真核生物中的作用是不对应的？

A. EF-G; eEF2

B. EF-Tu; eEF1α

C. RF1 和 RF3; eRF

D. EF-Ts; eEFαβ

7. 以下关于翻译后事件的陈述哪一个是错误的？

A. 一些 mRNA 编码多蛋白

B. 蛋白靶向与新生肽链的信号序列有关

C. 信号肽酶从一些蛋白的氨基末端去除一至两个氨基酸

D. 蛋白能被乙酰、磷酸化和糖基化等作用所修饰

R 噬菌体与真核生物病毒

1. 以下关于病毒的陈述哪一个是错误的?
   A. 病毒只能在宿主细胞内复制
   B. 一些病毒的被膜中含有宿主细胞蛋白
   C. 病毒基因组可以是双链或单链 DNA 或 RNA
   D. 复制缺陷型病毒可以通过互补作用来复制
   E. 所有的病毒都依赖于宿主细胞的复制和转录机制
   F. 有些病毒利用病症来辅助它们在宿主间的传播
2. 以下关于 M13 噬菌体的陈述哪一个是正确的?
   A. M13 噬菌体有双链 DNA 基因组
   B. M13 噬菌体通过与性伞毛结合而进入大肠杆菌宿主细胞内
   C. 多拷贝的复制型(RF)M13 是由正常的 RNA 引发的双链 DNA 复制而产生的
   D. 不同的 M13 噬菌体颗粒中 DNA 的数量差别很大
3. 以下关于 λ 噬菌体的陈述哪三个是正确的?
   A. λ 噬菌体具有双链 DNA 基因组
   B. λ 噬菌体与大肠杆菌外膜上的受体结合,将病毒 DNA 注入细胞内
   C. 裂解型生命周期需要将 λ 噬菌体基因组整合到宿主细胞基因组中
   D. *N* 和 *cro* 基因的终止是依赖于 rho 的
   E. λ 阻抑蛋白是抑制裂解作用的
   F. 宿主细胞的应激有利于启动裂解周期
4. 以下关于 DNA 病毒的陈述哪一个是错误的?
   A. DNA 病毒和宿主细胞的启动子之间经常有着序列同源性
   B. SV40 基因组有重叠的可读框
   C. SV40 大 T-抗原调节病毒和宿主细胞的转录和复制
   D. DNA 病毒调节宿主细胞复制的能力可能会导致宿主细胞产生肿瘤
   E. SV40 的 VP1、VP2 和 VP3 蛋白对于该病毒的致瘤特性是很重要的
5. 以下对于疱疹病毒的陈述哪一个是正确的?
   A. 疱疹病毒与致瘤作用无关
   B. 疱疹病毒基因组只有一条链上有基因
   C. HSV-1 编码它自身的 DNA 聚合酶
   D. HSV-1 有一个复制原点
   E. 疱疹病毒的潜伏型感染需要病毒基因组的染色体整合
   F. 所有的疱疹病毒基因对于病毒的复制都是必需的
6. 以下对于反转录病毒的陈述哪一个是错误的?
   A. 依赖 RNA 的聚合酶的精确性没有依赖 DNA 聚合酶高
   B. 像反转录病毒一样,酵母 Ty 转座子编码一种逆转录酶
   C. 反转录病毒有单链 RNA 基因组
   D. HIV 的 Rev 蛋白调节病毒的转录

S 肿瘤病毒与癌基因

1. 以下哪一种陈述不支持癌症是一种遗传疾病的观点？
   A. 有些癌症是遗传的
   B. 有的癌细胞有异常染色体
   C. 癌症是由致癌剂引起的
   D. 许多致癌剂能引起突变
2. 以下关于 NIH－3T3 细胞转染法分离癌基因的陈述哪两个是错误的？
   A. 它比体内分析法更简单
   B. 它比体内分析法更快
   C. NIH－3TS 细胞是正常的，但很容易被转化进肿瘤细胞内
   D. NIH－3T3 细胞适于吸收和表达外源 DNA
   E. NIH－3T3 细胞是可以检测细胞型特异性的癌基因的干细胞
3. 以下哪个蛋白群不含有癌基因产物？
   A. 转录因子
   B. 细胞表面受体
   C. 蛋白激酶
   D. 脂酶
   E. 多肽激素
4. 以下关于抑癌基因的陈述哪一个是错误的？
   A. 抑癌基因是性遗传的
   B. 抑癌基因比癌基因少
   C. 成视网膜瘤基因是肿瘤抑制基因
   D. 肿瘤抑制基因通过突变消除其正常的活性可以成为癌基因
   E. 肿瘤抑制基因与某些家族性癌症有关
5. 以下哪三个能在理论上增强癌细胞的形成或存活？
   A. 使 *bcl*－2 基因家族成员之一失活
   B. 使 *bax* 基因家族成员之一失活
   C. 过量表达 *P53* 基因
   D. 增加细胞内 BCL－2 蛋白与 BAX 蛋白的比例
   E. 去除 P53 蛋白
   F. 去除生长因子
6. 以下哪两个陈述是正确的？
   A. 凋亡是多细胞生物中细胞死亡的惟一机制
   B. 在所有的发育期，细胞凋亡的速率必须与细胞分裂的速率相同
   C. 凋亡是许多细胞在缺乏生长信号时默认的途径
   D. 凋亡需要对细胞原生质膜完整性的破坏
   E. 凋亡包括核 DNA 降解

答　案

A

1C，2B，3D，4B

B

1A，2D，3C，4A, 5D

C

1D，2C，3B，4B，5C

D

1D，2D，3C，4A，5C

E

1D，2C，3A，4D，5A，6A，7C

F

1D，2B，3A，4B，5C

G

1B，2D，3A，4A，5D

H

1D，2B，3C，4C，5D

I

1A/E，2A，3D

J

1A，2A/C/E，3E，4D，5C/D，6A，7B

K

1B/D，2B，3C，4E，5B

L

1A/E，2D，3D，4B/F，5D/F

M

1F，2A/F，3A，4B/E，5C，6D

N

1D/E，2B/F

O

1D，2C，3B，4C

P

1A，2C，3B/D/E

Q

1C，2B，3A，4E，5A/B，6D，7C

R

1E，2E，3A/B/F，4E，5C，6D

S

1C, 2C 和 E，3D 和 E，4A，5B 和 D 和 E，6C 和 E

索　引

2′-脱氧核糖　32
2′-氧甲基核糖　235
5-甲基胞嘧啶　59
7-甲基鸟苷　243
Acanthamoeba　210
anti 构象　36
AP-1　300
A-DNA　35～37
bcl-2　308
B-DNA　35, 36, 37
bZIP 蛋白　224
Caspases　309
CCAAT 框　216, 289
CDK　81, 220, 230
cDNA 文库　103, 139, 142, 147
CHEF　129
DNA 半糖基化酶　95
DNA 病毒
　非洲淋巴瘤病毒　214, 288
　基因组　287
　疱疹病毒　287～290
　乳多空病毒　287
　腺病毒　287, 290
　猿猴病毒 40　288
DNA 超螺旋　45～48
　缠绕　46～47
　连接数　45
　连接数的变化　46, 50
　扭曲　46～47
　扭转应力　47
　受约束的和不受约束的　50
　松散时闭环分子的连接数　46
　拓扑异构体　46
　在核小体中　52
　在琼脂糖凝胶中　111～112
　在原核生物中　52
　在真核生物中　52
　正的和负的　46, 185
　转录泡　188
DNA 重组
　Holliday 中间体　97
　不正常的　99
　普遍的　97
　同源的　97, 134
　位点特异性的　98
　在 DNA 修复中　98
DNA 促旋酶　48, 76, 188
DNA 复制　66
　2μ 复制起点　134
　DnaA 蛋白　76
　DnaB 蛋白　76
　RNA 引物　74, 76, 78
　半保留的　71
　半不连续的　74, 289
　病毒的　72, 83, 281
　常染色质　84
　单链结合蛋白　76
　端粒　57, 58, 85
　端粒酶　58, 85
　非洲爪蟾　84
　复制叉　71, 73, 74, 84
　复制起点，原点　72, 74, 84
　复制子　72
　冈崎片段　74
　滚动环　284
　后随链　74
　解链的，解旋的　76
　连环体　289
　模板　71
　酿酒酵母　83
　起点识别复合体　84
　起始　72, 74, 76
　前导链　74
　噬菌体　75, 286
　双向的　73
　脱氧核苷三磷酸　72
　校正　77, 88
　许可因子　84
　异染色质　84
　猿猴病毒 40　83
　增殖细胞核抗体　84
　质粒复制起点　105

忠实性 74,77,88
终止 72,73,77
自主复制序列 84,132
oriC 76
DNA 解旋酶 76
DNA 聚合酶 154
αDNA 聚合酶 84
δDNA 聚合酶 84
εDNA 聚合酶 84
ζDNA 聚合酶 90
聚合酶Ⅰ 77,95,99,104,154
聚合酶Ⅲ 77,84
DNA 克隆 103,116
cDNA 166～167
插入方向 117,152
基因的极性 167
基因组的 167
鉴定 146
鉴定 151
图谱 152,167
诱变 170～173
组织 166
DNA 克隆操作
DNA 克隆的应用 174～177
DNA 片段的分离 112
lacZ 基因 120
lacZ′基因 121
M13 的复制形式(RF) 126
T-DNA 134
Ti 质粒 134
X-半乳糖苷 120
β-半乳糖苷酶 120
λ 包装 125,129
λ 溶原菌(体) 122,126
包装提取液 125
表达载体 122
重组 DNA 114
穿梭载体 134
电穿孔法 134
多克隆位点(MCS) 121
酚-氯仿 106
酚抽提 106
辅助噬菌体 127
甘油贮液 117
感受态细胞 115
基因枪 134
碱裂解 106
菌落 116,147
抗生素抗性 105,119
蓝-白颜色筛选 120～121,126
连接 109,114
连接产物 119
黏性末端 109
黏性末端 109
片段的方向 117,152
染色体步移 148
染色体跳查 149
筛选 102,103,116,120
筛选标记 102
噬菌斑 126
双抗生素抗性 119～120
双酶切 117
宿主生物 102,105,115
微阵列 160
位置的 148～149
显微注射 134
限制酶消化 110
限制性酶切片段 109
小量制备 106
芯片 160
亚克隆 102
乙醇沉淀 106
异丙基-β-D-硫代半乳糖苷 120
影印平板培养 120
载体，见克隆载体
在酵母中筛选 132
在真核生物中 133～137
直接 DNA 转移 136
转化 103,115
转化效率 116
转染 134
组氨酸标记 122
DNA 连接酶 74,77,95,99,104,114,145
DNA 酶Ⅰ足迹法 168
DNA 缺陷的修复 96
DNA 损伤
3-甲基腺嘌呤 92
7-甲基鸟嘌呤 92
胞嘧啶脱氨基作用 91
苯并芘加合物 92
除嘌呤位点 92,95
嘧啶二聚体 88,92,94,95

脱嘌呤作用　92
烷基化　92
氧化的　92
O^6-甲基鸟嘌呤　92,94
DNA 拓扑异构酶　47～48
Ⅰ型　48
Ⅱ型　48,76
DNA 促旋酶　48,76,188
拓扑异构酶Ⅳ　48,78
DNA 修复
错配　88,95
光复活　94
核苷酸切除修复　95
碱基切除修复　95
适应反应　94
烷基转移酶　94
有错误倾向的　90
DNA 修饰
4-*N*-甲基胞嘧啶　35,109
5-甲基胞嘧啶　35,59,109
7-甲基腺嘌呤　35,109
DNA 引发酶　76
DNA 指纹法　64,175～176
E2F　81～82
erbB 基因　298
FIGE　129
fos 基因　300,303
F 因子　283
GTP 酶　299
G 蛋白　299
H-NS 蛋白　50
HU 蛋白　50
ICE 蛋白酶　309
Klenow 聚合酶　140,144,145,158
LINES　63
myc 基因　301,303
NIH-3T3 细胞
Northern 印迹　154～155
PFGE　128
Pribnow 框　185
RBS　67
RB 蛋白　82
RecA 蛋白　90,97
Rev 蛋白　294
RFLP　64
Rho 蛋白　181,191,198
RNA 编辑　67,250
RNA 病毒
HIV　66
反转录病毒　292
脊髓灰质炎病毒　280
RNA 复制　67
RNA 加工　234
RNA 聚合酶　67
T3 噬菌体　182
T7 噬菌体　182
α 亚基　183
β′亚基　183,206
β 亚基　183,206
σ 因子　182,187,201～203
大肠杆菌　179,182～183,187,201～203
核心酶　183,187
全酶　182,188
RNA 聚合酶Ⅰ　206,207～210,234
启动子　208
RNA 聚合酶Ⅱ　67,206,218
CTD　206,220,230
基础转录因子　218～220
启动子　215,289
增强子　216
RNA 聚合酶Ⅲ　206,211～214
启动子　214
终止　214
RNA 酶 A　104,106
RNA 酶 H　104
RNA 酶保护　236
RNA 转录图谱　168
rRNA　233～238
16S　234
18S　235
23S　234
28S　235
5S　234
5S 启动子　214
5S 转录　206,213
5.8S　235
反转录 PCR　164
基因　207
加工　234～236
甲基化　235
转录单元　189
转录起始　189

自我剪接 241
S-腺苷甲硫氨酸 234
S1 核酸酶 104,145,171
SD 序列 262
SINES 63
SNP 64
Southern 杂交 154
SP1 216,228
SPO1 202
SRP 受体 276
SSCP 65
SSLP 64
STAT 蛋白 229
T4 DNA 连接酶 104,141,145
T7 DNA 聚合酶 158
TAF_I 210
TAF_{II} 219
Tat 蛋白 230,285,294
TATA 框 215,289
TBP 210,212,215,219
TIF-1 210
tRNA 加工
原核的 240
真核的 239～240
Ty 反转录转座子 100,292
U6 小核 RNA 214
Wilms 瘤基因 226
X-半乳糖 120,168
X 射线衍射 236
X 衍射晶体学
蛋白质 29
核小体 52
Zoo 印迹 155
α-淀粉酶基因 249
α-鹅膏蕈碱 206
β-半乳糖苷酶 120,145,194
β-内酰胺酶 105
λ 噬菌体,见噬菌体 283～286
σ 因子 182,187,201～203
癌 281,288,293,295,302,309
癌基因 295～305
核的 300
种类 299
艾滋病病毒 159,294
氨苄青霉素 105,120～121
氨基酸 15～17
氨基肽酶 277
氨酰基-tRNA 252,258
转移至核糖体 270
氨酰基-tRNA 合成酶 258
氨酰基腺苷酸 258
摆动 261～262
孢子形成 202
报道基因 168～169
闭环 DNA 45,49～50
鞭毛 2,11
变性,失活 161
DNA 变性 39～40,44,154,161
RNA 变性 40,44
蛋白质变性 21
遍在蛋白 277
表达筛选 147
表达文库 147
病毒
包被 280
出芽 280
毒力 281
反受体 280
感染 281,289
核衣壳 279
基质蛋白 280
疱疹病毒 281
受体 280
衣壳 279
增强子 300
病毒基因组 280
病毒类型
单纯疱疹病毒-1 288～290
反转录病毒 281
非洲淋巴瘤病毒 288
感冒 281
互补 281,294
狂犬病 281
禽类白血病病毒 293
乳多空病毒 281
水痘-带状疱疹病毒 288
乙肝 281
猿猴病毒 40 281
病毒体 279
部分酶解 140,152
操纵基因序列 193
操纵子 67,193

测序
Maxam 和 Gilbert 156
Sanger 法 158
核糖核酸 158
化学法 158
酶法 158
脱氧核糖核酸 157～158
循环 165
自动的 159
缠绕 46
常染色质 58,84
超螺旋 DNA
沉降系数 5,12
成视网膜细胞瘤 303
成纤维细胞 230,296
程序性细胞死亡 307
持家基因 59,228
重叠基因 254
重复 DNA,分散重复的 100
重组,见 DNA 重组
重组 DNA 101
重组蛋白 27,122,174
重组动力学 62
触角足 222,225,231,243
串联基因簇 63
催化性 RNA 241
单核苷酸多态性(SNP) 64
单链构象多态性(SSCP) 65
单顺反子 mRNA 67
单序列长度多态性(SSLP) 64
蛋白测序 27～28
蛋白纯化 26～27,122,174
蛋白分泌 276
蛋白合成 66
30S 起始复合体 266
70S 起始复合体 266
80S 起始复合体 273
核糖体 67
30S 亚基 236
50S 亚基 236
A 和 P 位点 266
RNA 组分 11
蛋白质组分 11
结构特点 237
肽键形成 266
原核的 12,237
真核的 12,237
核糖体结合位点(RBS) 67
机制 265～270
起始 266,272
起始因子 266～272
扫描 271
释放因子 270
延伸 266
延伸因子 266
移码 293
真核起始 272
真核因子 272
终止 270
蛋白聚糖 9
蛋白酶复合体 278
蛋白酶体 278
蛋白酶消化 140
蛋白质
C 端 19
N 端 19
X-衍射晶体学 29
α-螺旋 20
β-折叠 20
超二级结构 23
等电点 26
二级结构 20
非共价作用 13,21,22
共轭的 20
核磁共振谱 29
基序 22
家族 22
结构域 13
氢键 13,20,21
球形的 19
三股螺旋 20
三级结构 20
疏水力 14,20,21
四级结构 21
羧基末端 19
肽键 19
纤维状的 19
一级结构 19
质量测定 28
质谱分析法 28
蛋白质组学 30
等密度离心 6,41,71,106

第二信使 299
电穿孔 134
电极 13,39
电泳
  CHEF 129
  FIGE 129
  PFGE 128
  聚丙烯酰胺凝胶 28
  琼脂糖凝胶 103,110～112,117～118
  双向电泳 30
淀粉 7
调控基因 193
定位克隆 148
端粒 57,58
端粒酶 9,58
多蛋白、聚蛋白 276
多核苷酸激酶 104,152
多核苷酸磷酸化酶 252
多核糖体 263
多聚 A 尾巴 67,143,244
多聚嘧啶序列 245
多克隆位点(MCS) 121
多联体 289
多顺反子 mRNA 67,194,275
多糖 7
  黏多糖 8
  糖基化 5
恶性肿瘤 295
二氢尿嘧啶核苷 256
发夹结构 181,198
翻译调控 275～276
翻译后修饰 15
翻译体系,无细胞的 143
翻译移位 293
反密码子 252,256
反密码子脱氨酶 261
反向重复 194
反义 RNA 276
反义链 154,178
反终止子 200
反转录 DNA 103,139,142～145
反转录病毒 66,100,292
  gag 基因 293
  HIV 66
  反转录酶 67,281,292
  禽类白血病病毒 293
  突变率 292,294
  移码 293
  原病毒 292
  整合酶 292
  env 基因 292
  pol 基因 292
反转录酶-PCR 164
反转录酶 67,104,143,281,292
反转录转座子 100,292
范德华力 13,21
放射自显影 147
非洲淋巴瘤病毒 214,288
分化 2,3,230
分支点序列 245
分子伴侣 12,21
酚-氯仿 106,140,152
辅基 19,21
复制体 77
复制型 126,283
干扰素 174,229
甘油贮液 117
肝素 183
高尔基复合体 3,5
功能基因组学 160
共有序列 185
共阻抑物 198
古细菌 2
寡聚 dG 145
寡聚 dT 143
寡聚核苷酸 146
  接头 145
  引物 154,158,161,167,172
果蝇 3,159,231
  P 元件转座酶 250
合成三核苷酸 252
核 3,4
  核孔 3,4
  核膜 3,4
核癌基因 300
核蛋白 9,12
核定位信号(NLS) 276
核苷 32
核苷酸 32
  加成 234,277
  去除 234,240
  修饰 234,240

anti-构象　36
syn-构象　36
核基质　54,57,84
核酶　9,236,241
核内不均一 RNA,见 hnRNA
核内不均一 RNA　243
核内不均一 RNP　243
核仁　3,4
核酸
3′端　32
5′端　32
变性　40,43
除嘌呤的　39
定量　43
核苷　32
核苷酸　32
减色性　43
碱基　32
碱基互变异构体　39
碱效应　39
均匀标记　152
链特异性标记　153
末端标记　152
酸效应　39
探针　146
糖-磷酸主链　32
退火　44
稳定性　38
杂交　44
紫外吸收　42
最大吸收波长　43
核酸酶　52,239
DNA 酶Ⅰ　58,154,168
RNA 酶Ⅲ　234
RNA 酶 A　104,106
RNA 酶 D　239
RNA 酶 E　239
RNA 酶 F　239
RNA 酶 H　104
RNA 酶 M16　234
RNA 酶 M23　234
RNA 酶 M5　234
RNA 酶 P　239
RNA 酶 Phy M　158
RNA 酶 T1　158
RNA 酶 U2　158
S1　104,167
单链的　145
枯草杆菌核酸酶　158
绿豆核酸酶　104,145,171
内切核酸酶　52,239
外切核酸酶　52,77,88,239,243,276
外切核酸酶Ⅲ　104,170
微球菌核酸酶　52
限制酶　108
核酸酶 S1　167
核糖　32
核糖核蛋白(RNP)　58,234,240
电子显微镜术　236
交联　236
抗体　236
脱离,释放　236
再装配　236
核糖核蛋白
核糖核酸
测序　158
成熟　234
发夹结构　188,199
反义链　154
复制　67
合成　179～181
碱基　32
碱水解　40
结构　31～37
结合试验　236
茎-环结构　230
链起始　188
链终止　188
氢键　34
酸效应　39
稳定性　38～39
修饰　37
延伸　188
有义链　154
紫外吸收　42
核糖体　67
核糖体 DNA　63
核糖体蛋白　236
核糖体结合位点　67,122,262,276
核糖体受体蛋白　276
核小体　12,52,58,59,84
核衣壳　279

核移植 175
互补作用 281,294
环腺苷酸受体蛋白 185,195～196,198
回文结构 194,198
肌动蛋白 2,11,22
肌钙蛋白 T mRNA 250
肌苷 256,261
肌球蛋白 12,22
肌球蛋白 D 82,224,230
肌球蛋白基因 249
基因表达 67
基因剔除 175
基因文库 103,139～149
  大小 140
  代表的 139
  反转录 DNA 103,139
  基因组的 103,139
  筛选 146～149
基因治疗 102,177
基因组
  病毒 281
  大肠杆菌 2,49,62,159
  果蝇 159
  甲烷球菌 2
  酿酒酵母 159
  人 159
  细菌支原体 2
  线虫 159
  真核的 62
基因组测序 159
基因组文库 103,139
基因组学 159
  DNA 微阵列 160
  DNA 芯片 160
  功能基因组学 160
脊髓灰质炎病毒 279
甲基化 37,58,92,95,108,156,246,277
甲基甲磺酸 89,93
甲酸 156
甲酰胺 40,155
甲状腺激素受体 226,229,301
假尿苷 256
兼性离子 15
减数分裂 97
剪接 68,245
  snRNP 68
  变化的 248～249
  内含子 68
  外显子 68
  自身 241
剪接体 .246
碱基互变异构体 88
碱性磷酸酶 104,113,115,124,145,154
鉴别元件 258
胶原 20,22,277
角蛋白 12,22
酵母人工染色体 131
  Z-DNA 35,36,37
  插入 159
  载体 141
结构基因 193
经过遗传改造的生物(GMO) 175
茎-环结构 181,188,199
肼 156
聚合酶链反应(PCR)
  不对称的 165
  定量的 165
  多聚体 164
  反向 PCR 164
  反转录(RT-PCR) 164
  简并寡核苷酸引物(DOP) 164
  快速扩示 cDNA 末端(RACE) 164
  酶 164
  镁离子浓度 164
  模板 162
  嵌套的 164
  退火温度 162,164
  引物 162,172
  优化 164
  诱变 165,172
  周期 162
聚腺苷酸合成酶 244
聚腺苷酸化 143,236
  可变加工 248
聚乙二醇 152
卷曲螺旋结构 224
菌落 147
菌落刺激因子 1(CSF－1) 298
菌落杂交 147
菌毛伞毛 2,283
抗体 22,147
  多样性 98

抗原决定基 147
可读框 151,254
可读框 254
克隆,见 DNA 克隆
克隆化,见 DNA 克隆化
克隆载体 102,141
  2μ 质粒 134~135
  M13 102,126~127
  pBR322 119~120
  Ti 质粒 102,134~135
  λgt11 145,147
  λ 载体 102,123~126
  λ 置换质粒 124
  表达载体 122,287
  病毒的 136
  哺乳动物的 287
  穿梭载体 134
  反转录病毒 102,136
  杆状病毒 102,136
  酵母人工染色体(YAC) 102,131~132
  酵母游离质粒(YEp) 102,134
  黏粒 102,129~130
  噬菌体 102,123~127
  细菌人工染色体(BAC) 102,132,159
  猿猴病毒 40 102,136
  杂合质粒 M13 127
  质粒 102,105,119~122
枯草杆菌 202
类固醇激素 228
  受体 228
  效应元件 229
  终止密码子 67,252,270
类组蛋白 50
离心
  差速离心 5
  等密度离心 6,41,71,106
  速度区带离心 6
利迪链菌素 182,183
利福平 183
利什曼原虫属 250
连接数 45
亮氨酸拉链 224
裂解性生命周期 123,284
磷酸二酯键 33
磷酸化 277
硫酸二甲酯 156
绿豆核酸酶 104,145,171
氯化铯 41,71,106
螺旋-环-螺旋结构域 224,231
螺旋-转角-螺旋结构域 222~223,231
麦芽提取物 143
慢病毒属 293
密度梯度离心 6,41,71,106
密码子-反密码子相互作用 261
密码子 252,256
  同义的 253
嘧啶 32,253,256
免疫球蛋白 250,300
模板链 179
膜
  核膜 4
  膜蛋白 13
  膜结构 13
  内质网膜 5
  原生质膜 1~3
末端标记 152
末端转移酶 104,145,154
内部转录间隙(ITS) 235
内含子 62,68,240,245
内质网 3,5,276
尼龙膜 147,154
拟核,类核 1,49
黏蛋白 9
黏粒载体 102,129~131,140,141
黏性位点(末端) 124,284
鸟嘌呤 32
尿嘧啶 32
尿素 40
凝胶阻滞(凝胶迁移) 168
扭曲(w_r) 46
哌啶 156
疱疹病毒 281
  HSV-1 病毒 288~290
嘌呤 32,253,256
启动子 67
  -10 序列 185,188,201~202
  -35 序列 185,188,201~202
  RNA 聚合酶Ⅰ 208
  RNA 聚合酶Ⅱ 215
  RNA 聚合酶Ⅲ 211
  大肠杆菌 179,182~186,188,202
  热休克 202

起始 tRNA 263,266,272
起始密码子 67,266
起始元件 216,220
前 mRNA 67
潜伏期 289
嵌入剂 47,89
羟基磷灰石层析 62
切口平移 154
禽类白血病病毒 293
氢键 13,17,22
琼脂糖 110~112
染色单体 57
染色体
DNA 环 49~50,57
DNA 结构域 49~50,54,57
超螺旋 49,52~53
大肠杆菌 49
动粒 57
端粒 57,58,131
纺锤体 57
分裂间期 58
核基质 54,57,84
畸形 295
微管 57
细菌人工染色体(BAC) 132,159
性染色体 X 58
有丝分裂的 57
原核的 49~50
真核的 54,56~60
支架 57
着丝粒 57,131
染色体步移 148~149
染色体跳查 149
染色小体 53
染色质 12,51~55
30 纳米纤维 54,57,58,59
CpG 甲基化 58~59,141
DNA 酶 I 敏感性 58
常染色质 58,84
高度有序结构 54
核基质 54,57,84
核小体 52~53
螺线管 54
染色小体 53
衔接 DNA 53
异染色质 58,84
组蛋白 12,52~54,60,277
热休克
蛋白质 198
基因 168
启动子 185
Helicobacter pylori 159
人免疫缺陷病毒 66,292
Rev 蛋白 294
TAR 序列 230
Tat 蛋白 230,285,294
基因表达 230
人生长激素 102
溶原性生命周期 123,124,284
乳多空病毒 287
乳糖 194
乳糖诱导物 195
弱化作用 198~200
扫描假说 272
色氨酸
操纵子 198
前导 RNA 199
前导肽 199
弱化子 198
阻抑物 198
上游 179
上游调节元件(URE) 216,222
上游结合因子(UBF) 208
上游控制元件(UCE) 208
生长因子 298,299
生物信息学 159
十二烷基硫酸钠 106,147
示踪标记 143
释放因子 270,274
噬菌斑 126,147
噬菌斑复制 147
噬菌斑杂交 147
噬菌体
M13 噬菌体 102,126~127,279,283
Mu 噬菌体 283,286
RNA 聚合酶 182
SP6 121
SPO1 202
T4 噬菌体 75
T7 噬菌体 75,121,122
λ 噬菌体 98,102,123~126,159,283~286
σ 因子 202

辅助噬菌体　127
噬菌体包装　125,141
噬菌体纯化　151
噬菌体繁殖　125,141
噬菌体整合　282
噬菌体转座　282
ΦX174 噬菌体　75,254
噬菌体
受体,细胞表面　299
疏水性　14,17,22,26
疏水作用　39
数据库　158
双脱氧核苷酸　158
水痘-带状疱疹病毒　288
顺反子　243
四环素　105,120
羧基末端区　206,220,230
羧基肽酶　277
肽基转移酶　270
糖苷键　32
糖基化　5,277
糖原　8
糖脂　10
套索　245
体外转录　147
停靠蛋白　276
同聚物,均聚体　145
同义密码子　253
同源框　222,231
同源异型基因　231
同源域　222,231
突变　295,304
错义　88
颠换　87,253
点　87,90
无义　88
移码　88
隐性　302
转换　87,253
自发　88,91
突变剂
芳基化剂　89,92
放射性　88
碱基类似物　89
嵌入剂　89
烷化剂　89,92
亚硝酸　89
图式形成　231
脱磷酸作用　145,154
脱氧核糖核酸(DNA)
A-DNA　35～37
A_{260}/A_{280}比值　43
B-DNA　34,36～37
CpG 甲基化　58～59,141
DNA 结合蛋白　50,222～226
G+C 含量　41,63
SINES　63
Z-DNA　36,37
闭环的　45,49
变性　39～40,44,154,161
部分酶解　140,152
测序　104,156～160
缠绕　46～47
长散布元件　63
超螺旋,见 DNA 超螺旋
超声处理　40,62,140
纯度测定　43
大沟　34,35
定量　43
多变的　64
反向平行链　35
非编码的　62
分散重复的　63
浮力密度　41
复性　44
复杂度　62
高度重复的　62
核糖体的　63
互补链　34
甲基化　37,58,92,95,108,156,277
剪切　62,140
碱基　32
碱基对　34～35
碱基特异性的切割　156
碱效应　39～40
接头　145
解链,熔解　44
解链温度　44
开环的　111
连接　103,114,141,145
连接数　45
连接数的变化　46～47,49～50

末端标记 152
末端修复 140
黏性 40
扭曲 46
扭转应力 47
嵌入物 47,89
氢键 35～37,38
缺刻的 111
热变性 44
双螺旋 34～36
松弛的 46
松散时闭环分子的连接数 46
酸效应 39
探针 103
特有序列 62
退火 44,109
拓扑异构酶 47～48
弯曲 196
微卫星 63
卫星 58,63～64
稳定性 38～39
衔接子 145
线粒体的 4～5
小沟 34,35
小卫星 63
修饰 37
溴化乙锭结合 47,111
叶绿体 5
杂交 44,103
指纹法 64,175～176
中度重复的 62
重新缔合动力学 62
轴比 40
紫外吸收 42
自动合成 146
外部转录间隙 235
外显子 22,68,245
可变外显子 248
网织红细胞裂解物 143
微管 2,11,12
微管 2,11,22,57
微球菌核酸酶 52
微体 3
卫星 DNA 58,63～64
细胞
细胞分化 2,3,230
细胞决定 230
细胞转化 281
细胞壁 1
细胞凋亡 306～309
单纯疱疹病毒-1 288～290
反转录病毒 281
非洲淋巴瘤病毒 288
互补 281,294
狂犬病 281
乳多空病毒 281
水痘-带状疱疹病毒 288
乙肝 281
猿猴病毒 40 281
细胞分级分离 6,140
细胞骨架 2,3,11,22
微管 2
微丝 2,11
中间纤维 11
细胞质 1,2
细胞周期 79～82,220,230
DNA 合成期 80
E2F 81～82
G_0 期 80
G_1 期 80
G_2 期 80
M 期 80
S 期 80
成视网膜细胞瘤基因(*RB*) 81～82
促细胞分裂源 80
分裂后期 80
分裂间期 80
分裂前期 80
分裂中期 80
关卡 80～81
激活 82
间期 80
静止期 80
限制点(R 点) 80
抑制 82
有丝分裂 80
周期 80
细胞周期蛋白 81
细菌人工染色体(BAC) 102,132,159
纤毛 11
纤维素 7
显性负效应 304

线虫　159,307
线粒体　3,4
限制酶　108,140
　部分消化　152
　回文序列　109
　黏性末端　109
　黏性末端　109
　识别序列　109
　双酶切　117,152
　作图　151～152
　*Bam*HⅠ　140
　*Eco*RⅠ　109
　*Sau*3A 内切核酸酶　140
限制性酶切图谱　103,149,151～152
限制性片段长度多态性(RFLP)　64
腺嘌呤　32
相抽提　106,140
硝酸纤维素膜　147
小核 RNA　243
　转录　206
小核核糖核蛋白　68,235,241,243
小鼠乳腺瘤病毒(MMTV)　297
锌指结构域　224
信号识别颗粒(SRP)　276
信号肽　276
信号肽酶　276
信号序列　276
信号转录　229
信使 RNA
　poly(A)尾　68
　大小　143,155
　单顺反子的　67,275
　多顺反子的　67,194,275
　分级　143
　分离　143
　富集,浓缩　143
　合成的　252
　加工　242～247
　加帽　67,243
　甲基化　246
　降解　243
　聚腺苷酸化　244
　可变加工　248～250
　拼接,剪接　245～246
　起始密码子　67
　隐蔽的　276
　终止密码子　67
　珠蛋白　143
胸腺嘧啶　32
胸腺嘧啶　32
胸腺嘧啶核糖核　256
溴化乙锭　47,111
溴脱氧尿苷　85
序列数据库　158
血红蛋白　21
血小板衍生生长因子　299
血友病　174
亚克隆　103
叶绿体　4～5
依赖细胞周期蛋白的激酶(CDK)　81,220,230
胰岛素　102
移码　293
移位酶　270
遗传多态性　63～64,88
遗传工程　102
遗传密码　66,251～254
　表　253
　简并性　252
　解读、破译　252
　特征　253
　通用性　254
　同义密码子　253
　突变　253
　修饰　254
乙醇沉淀　106
乙基亚硝基脲　89,92
异丙基-β-D-硫代半乳糖苷　120,195
异染色质　58,84
异乳糖　194～195
抑癌基因　292,302～305
易位　266,270
因子Ⅷ　174
引物　154,162,172
　简并的　162
引物延伸　167
隐性突变　295
印迹
　Northern 印迹　154～155
　Southern 印迹　154～155
　Zoo 印迹　155
影印平板培养　147
用于 DNA 克隆的酶　104

有丝分裂 57,80
有义链 154,179
诱变
定向位点 172
间接 89,91
聚合酶链反应 172
缺失 170～171
直接 89,91
转移损伤 DNA 合成 89
原癌基因 295
原病毒 292,296
原核生物 1
原生质体 134
原噬菌体 284
猿猴病毒 40 83,216,288
杂合细胞 302
杂交 44,146
菌落 147
噬菌斑 147
探针 297
严紧性 155
杂交分子捕获翻译 148
杂交分子释放翻译 148
载脂蛋白 250
增强子 216,300
蔗糖梯度 143
真核生物 1
真细菌 1,2
整合酶 98
整合宿主因子 50
脂蛋白 10,22
脂类 7,10
指导 RNA 250
质粒载体,见克隆载体
致癌病毒
DNA 病毒 288
反转录病毒 293,296
致癌作用 89,92,295
中心法则 66
中子衍射 236
终产物抑制 198
终止序列 181,188
肿瘤病毒 295～305
肿瘤形成 281,288,293
珠蛋白 mRNA 143
转化子 117
保存 117
筛选 117
转基因生物 175
转甲酰酶 263
转录 66
闭合复合体 188
单元 180
调控目标 226
复合物 180,226
开放复合体 188
启动子 67
启动子清除 188
起始 179
起始位点 167,180,185
三元复合物 188
体外 154
延伸 181,188
终止 67,181
终止信号 188
终止序列 181,188
转录泡 188
阻抑物区 225
转录因子 22,221～226
DNA 结合区 222
RNA 聚合酶Ⅰ 207～210
RNA 聚合酶Ⅱ 218
RNA 聚合酶Ⅲ 211～214
SP1 216,228
TBP 210,212,215,219
激活区 225
结构域 222
结合位点 168,171
磷酸化 277
域互换试验 222
转录体 30
转染 134,297
转移 RNA 67
CCA 末端 240
D 环 212,256
T 环 212,257
氨基酸接受臂 256
摆动 261～262
被修饰的碱基 255
不变核苷酸 256
二级结构 256
功能 258

核苷酸转移酶　240
基因　212
加工　239～240
结构与功能　255
可变臂　256
起始密码子　263,272
三级结构　257～258
三叶草　256
校正　259
一级结构　255
转录　206,211
转座　63
L1 元件　63
Tn 转座子　99
插入序列　99
反转录转座子　100,292
噬菌体　286
Alu 元件　63
Ty 元件　100,292
着丝粒　57
紫外辐射　147
自身免疫病　281
自我剪接　241
组氨酸操纵子　200
组蛋白　12,52,84,277
H1　52,53
H5　53,60
八聚体核心　52
变异体　60
核心　52
磷酸化　60
乙酰化　60
Alu 元件　63
EcoR Ⅰ 甲基化酶　145
erbA 基因　301
fms 基因　299
int -2 基因　297
jun 基因　301
lacZ 基因　120,147,194
L1 元件　63
N -甲酰甲硫氨酶　263,271
P53 基因　304
ras 基因　299
RB1 基因　303～304
sis 基因　299
Syn -构象　36
Taq DNA 聚合酶　104
Tetrahymena thermophila　236
trpR 操纵子　198

收到期 2003. 7. 24.
来源 三联
书价 38.00元
单据号 000353
开票日期 2003. 7. 24.

中科院植物所图书馆
S0000193

000027522
58.178
487
2001.
分子生物学

借者单位	借者姓名	借出日期	还书日期

58.178
487
000027522

(Q-0961.0104)
责任编辑／丁海珈　谢灵玲·封面设计／黄华斌　高海英

生命科学编辑部
联系电话：010-64012501
http://spbio.yeah.net
e-mail:spbio@163.net

21世纪国际化新概念教材

新！
快！
全！
易！

现代生物学精要速览系列

英文影印版	中文翻译版
《遗传学》	《遗传学》
《生态学》	《生态学》
《免疫学》	《免疫学》
《心理学》	《生物化学》
《生物化学》	《微生物学》
《微生物学》	《分子生物学》
《分子生物学》	《神经科学》
《神经科学》	《有机化学》
《有机化学》	《无机化学》
《无机化学》	《物理化学》
《物理化学》	《动物生物学》
《动物生物学》	《生物学中的化学》
《发育生物学》	……
《生物学中的化学》	
……	

ISBN 7-03-008383-0
9 787030 083838 >

ISBN 7-03-008383-0
定　价：38.00 元